Dissertation zur Erlangung des Doktorgrades

der Fakultät für Chemie und Pharmazie

der Ludwig-Maximilians-Universität München

Reprogramming *Emesvirus zinderi*

–

An approach to changing an RNA virus's behaviour

von

Alexander Wagner

aus

München, Deutschland

2022

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von **Herrn Prof. Dr. Hannes Mutschler** betreut.
und von **Herrn Prof. Dr. Klaus Förstemann**
von der Fakultät für Chemie und Pharmazie vertreten.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, ...20.09.2022.....

Alexander Wagner
.....

Dissertation eingereicht am .....04.07.2022.....
1. Gutachter: Prof. Dr. Klaus Förstemann
2. Gutachter: Prof. Dr. Hannes Mutschler

Mündliche Prüfung am .....11.08.2022.....

Contents

Zusammenfassung

Die Genome aller heutigen Organismen sind in Desoxyribonukleinsäure (DNA) kodiert. In den frühen Stadien der Entstehung des Lebens könnten jedoch Ribonukleinsäuren (RNA) eine vergleichbare Rolle gespielt haben. Mögliche Nachfolger dieser ursprünglichen RNAs finden sich in Form von RNA-Viren, insbesondere kleinen einzelsträngigen RNA-Viren. Ein gut untersuchtes Beispiel für diese Viren ist *Emesvirus zinderi*, besser bekannt als Bakteriophage MS2. In dieser Studie diente das Genom von MS2 als Ausgangspunkt für den Entwurf eines RNA-basierten Replisoms, das für genetische Informationen über das Replikase-Gen hinaus kodiert. Ausserdem wurde unter Ausnutzung des Infektionsmechanismus von MS2, der auf dem MS2-Reifungsprotein basiert, parallel dazu versucht ein neues RNA-Transportsystem zu entwickeln. Diese Ansätze sind nicht nur für die synthetische Biologie und biochemische Technologien von Interesse, sondern könnten auch Aufschluss über die Rolle von RNA am Anfang des Lebens geben.

Durch den Einsatz klassischer biochemischer und mikrobiologischer Methoden in Kombination mit der modernen synthetischen Biologie konnte eine Vielzahl von Werkzeugen für dieses Vorhaben etabliert und umgenutzt werden. Der Einsatz eines in der synthetischen Biologie verwendeten *In-vitro*-Translationssystems führte zur Identifizierung von Aminosäuren im Reifungsprotein, die für die virale Infektion entscheidend sind, während klassische Methoden zur Proteinaufreinigung und zum Nachweis dieser Infektion die *In-vitro*-Rekonstitution von minimalen infektiösen Einheiten von MS2 ermöglichten. Unter Verwendung desselben Translationssystems führte die sequenzielle Reduktion/Vereinfachung desselben zur Identifizierung der essenziellen Proteinfaktoren für die RNA-Replikation und zusätzlicher regulierender Co-Faktoren. Darüber hinaus führte die Charakterisierung der Replikationsmaschinerie zur Entdeckung neuartiger RNA-Replikatoren, die eine Plattform für den Entwurf von selbstreplizierenden RNA-Vektoren bildeten. Diese Replisomen, die in ihrer Funktion den plasmidbasierten DNA-Replisomen ähneln, waren nachweislich in der Lage *in vitro* zu replizieren und translatiert zu werden. Schließlich wurden im Rahmen dieser Arbeit zusätzliche Methoden etabliert, die Potenzial für künftige Untersuchungen von RNA-Replisomen aufweisen, aber auch eine Strategie für die Erzeugung von Replisomen auf Basis anderer viraler Systeme bieten.

Zusammenfassend lässt sich sagen, dass das Ziel dieser Studie, die Entwicklung eines RNA-basierten Replisoms, erreicht wurde und diese Entdeckung neue Anwendungsmöglichkeiten in den Forschungsbereichen der synthetischen Biologie und der Entstehung des Lebens eröffnet.

Summary

The genomes of all of today's organisms are encoded in deoxyribonucleic acid (DNA). However, at the early stages of the emergence of life, ribonucleic acids (RNA) might have played a comparable role. Potential descendants of these ancient RNAs can be found in the form of RNA viruses, especially small single stranded RNA viruses. A well-studied example for these viruses is *Emesvirus zinderi*, more commonly known as bacteriophage MS2. In this study, the genome of MS2 served as a starting point for the design of an RNA based replisome, encoding for genetic information beyond the replicase gene. Furthermore, harnessing the infection mechanism of MS2, based on the MS2 maturation protein, it was tried in parallel to establish a new RNA delivery system. While these approaches are of interest for synthetic biology and biochemical technologies, they also might shed light on the role of RNA at the beginnings of life.

Using classical biochemical and microbiological methods in combination with modern synthetic biology, a variety of tools could be established and repurposed for this endeavour. Deploying an *in vitro* translational system, used in synthetic biology, led to the identification of amino acids in the maturation protein crucial for viral infection, while classical methods for protein purification and detection of this infection enabled the *in vitro* reconstitution of minimal infectious units of MS2. Using the same translational system, sequential reduction/simplification of it led to the identification of the minimal complex for RNA replication and regulating co-factors. In addition, the characterisation of the replication machinery led to the discovery of novel RNA replicators, providing a platform for the design of RNA replisomes. These replisomes, similar in function to plasmid-based DNA replisomes, could be shown to be capable of *in vitro* replication and translation. Finally, during this thesis additional methods were established that exhibited potential for future studies of RNA replisomes, but also provide a strategy for replisome generation based on alternative viral systems.

In conclusion, the goal of this study, the design of an RNA based replisome was achieved and this discovery opens new possible applications in the research fields of synthetic biology and the emergence of life.

Publications

This thesis is partially based (Sections 3.6 – 3.8) on results published in the following article:

Wagner A., Weise L. I., Mutschler H., *In vitro* characterisation of the MS2 RNA polymerase complex reveals host factors that modulate emesviral replicase activity, *Commun. Biol.* (**5**), 264 (2022)

In addition, parts of this work (Section 3.4) are based on results from the master thesis of *Alexander Floroni*, which was planned and supervised by the author.

Floroni A.; *'Quantification of electroporation-mediated uptake of RNA molecules into Escherichia coli'*, 2021

1. Introduction

1.1 Hereditary information

Within the human species, multiple types of information are passed on from generation to generation. These are biological, as well as social/cultural information [1]. The biological information is defined by the genes of an individual that are inherited from the parents, with half of the progeny genome stemming from each parent. In case of cultural information, numerous different forms exist, ranging from traditions, social norms, and most importantly language. Thereby, both genetic and linguistic information share distinct characteristics, which was already noted by *Darwin* in his work on the origin of species [1]. First, in both cases the information can be broken down into smaller bits. For language, these building blocks are words, that consist of letters and can be combined to sentences. For a genome, words would correspond to genes and their regulators, that are assembled from chemical building blocks called nucleotides. The genome is then the combination of all genes and genetic regulators of an organism [2]. For both types of information, the sequence of the individual building blocks is crucial, and permutations potentially change the meaning. Language serves the purpose to transcribe one's thoughts into a message, that can be interpreted by someone else. Here, the second individual translates the message using their understanding of the language and thus interprets it. The genes encoded for by a deoxyribonucleic acids (DNA) genome of an organism are transcribed by enzymes into the so-called messenger ribonucleic acid (mRNA), which in turn gets translated by large protein/RNA complexes, the ribosomes [3,4]. In the process, the mRNA is deciphered with the help of transfer RNA (tRNA), that enables the read-out of the nucleotide sequence into the respective amino acid sequence of the product protein. The thereby synthesised protein corresponds to the function that was encoded by the genetic information, like the intention of a spoken or written message for linguistic information. Furthermore, just like genomes, languages can evolve over time [1,5,6]. The process of genome replication and maintenance is error prone, thus miniscule changes can accumulate over time, slowly changing the originally encoded genetic information [5,6]. This process is driven by selection pressure on a population, where changes that provide a benefit are more frequently passed on than deleterious ones, until they make up the majority within a population [7]. In a similar fashion, languages evolve. Here, the miniscule changes are introduced by the everyday use and adaptions to a language, made by offspring generations [1]. This approximates the replication and maintenance of a genome. Just like genes, novel words that fill a need will quickly become a stable part of a language, while those that do not anymore will soon be lost. By comparison of the respective components of information – genome and language – it is possible to trace back the evolutionary origin (Figure 1A, B) [8–11]. In the shown example (Figure 1A), the branching off of new Germanic languages from a shared ancestral language – Proto-

Germanic – is depicted using a glossogeny tree. Here, phonetic, and etymological changes over time led to the separation into distinct languages [1]. The more diverse the individual languages are, the farther apart they are shown in the tree [8,9]. Noteworthy, an emerging language does not necessarily only have a single ancestor but can originate by crossing of closely related languages. Thus, Middle Upper German was derived from both Irminonic and Istvaeonic languages [12]. However, new languages can not only emerge, but old ones can also go extinct. In this case the East Germanic language family was lost in the process [12,13].

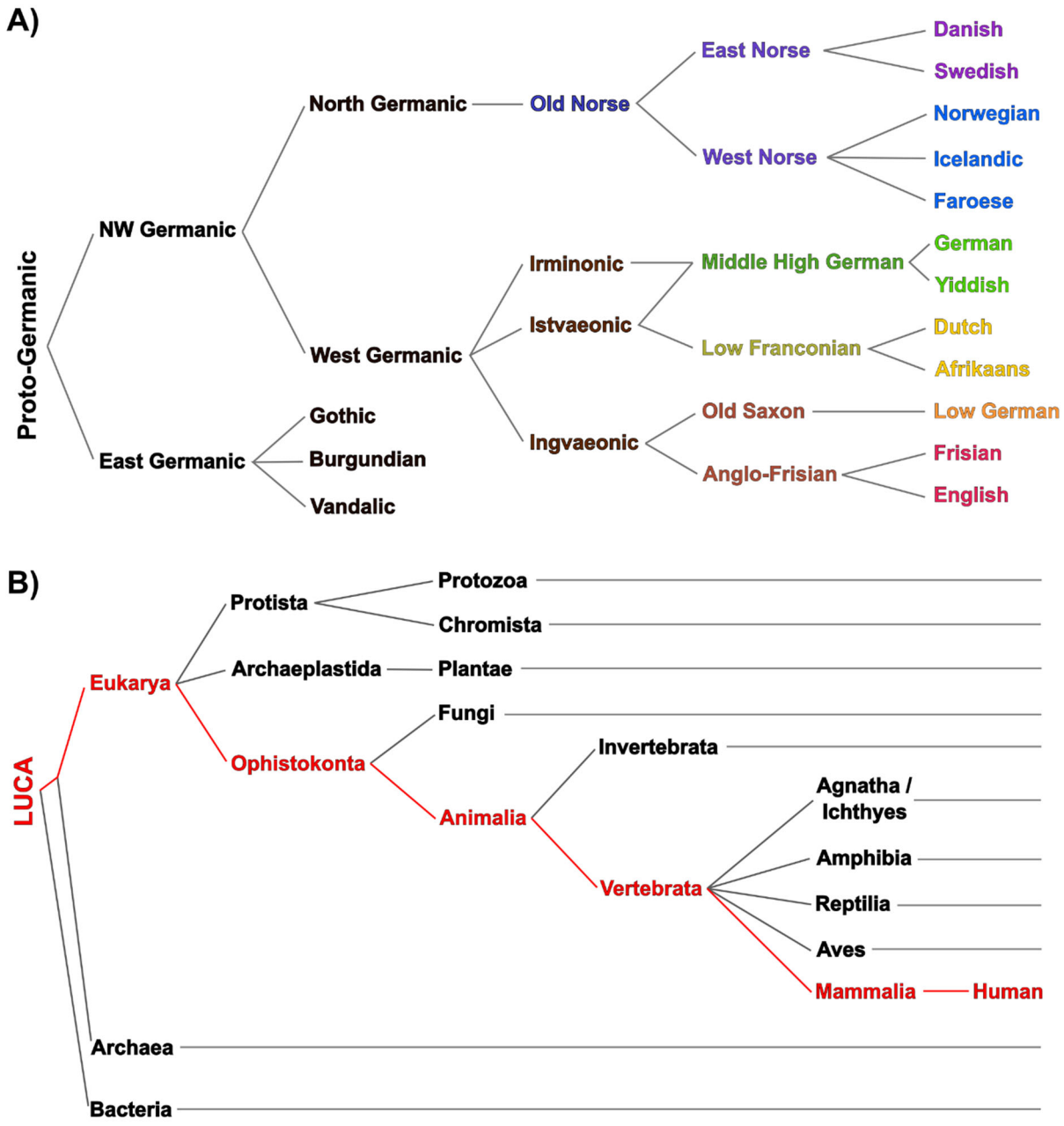


Figure 1 Phylogeny trees: A) Simplified phylogenetic tree depicting the evolution of Germanic languages, starting from an ancestral Proto-Germanic [12,13]. **B)** Simplified phylogenetic tree, showing the evolutionary lineage (red) connecting humans to the last universal common ancestor (LUCA) [14–16].

An analogous approach allows the identification of the ancestral origin of a species. Thus, a phylogeny tree can be made, that exhibits for example the evolution from the so-called last universal common/cellular ancestor (LUCA), at the beginning of cellular life, to present-day humans (Figure 1B) [14–16]. Based on the similarity in function, structure or sequence of genes or proteins, the chronological order and relation between different species can be established [10,11]. Identical to languages, external factors like environmental changes are thereby the main drivers for the emergence, but also extinction of species [17]. Using the example of the evolution of *Hominids*, several extinct archaic subspecies are known, including *Homo neanderthalensis* and the putative *Homo denisova* [18,19]. The evolution of humans is also a good example for the emergence of interspecific hybrids – the modern *Homo sapiens* – by introgression of *H. neanderthalensis* and *H. denisova* with archaic *H. sapiens* [18,19]. In general, however, the definition of what constitutes already a new species/language and what has still to be considered as a subspecies/dialect, is often not as sharp as depicted.

This is even more complicated when a set of information contains bits of information of another set, that is a genome containing genes of a different species, or a language containing words from a different one [1,20]. In case of genomes, these can derive from horizontal gene transfer (HGT). Here, genetic information is not passed on along the ancestral-progeny line (vertical gene transfer), but is rather acquired from outside the own species, for example because of viral infection or integration of other mobile genetic elements [20,21]. The linguistic equivalent to HGTs are loanwords (words that are directly adapted from a different language), calques (words that are adapted after literal translation from another language), or slang [22,23]. In addition, while genes are only inherited from the parents, the vocabulary of an individual is also shaped by its social environment [1]. Thus, children often will adapt words from peers, or strangers [1]. Finally, slang does not adhere to the strict rules of its respective formal language [23]. Similarly, viruses are notorious for bending the rules of cellular life [24]. A good example for this are internal ribosome entry sites (IRES). The canonical translation from mRNAs usually requires a distinct set of translation factors for initiation, elongation, termination, and factor recycling [4]. IRES, specific structural elements in the viral mRNA, though, enable translation initiation without the usually required initiation factors [4].

In contrast to languages, however, the flow of genetic information is unidirectional. A thought can be put into words, interpreted by the receiver and in turn put back into words to be interpreted again. Once the information encoded by DNA was transcribed into mRNA and translated into a protein, there is no way back from protein to DNA or RNA. This idea was first introduced by *Francis Crick* in 1957 and has since then held true [3]. Furthermore, the minimal building blocks of a language, the used letters, can greatly vary. Thus, the Japanese language knows three sets of alphabets, Hiragana, Katakana and Kanji, all different to each other, as well as to the Latin alphabet used by many European languages. The genetic information of

all living cells, however, is encoded by DNA, built from only four different deoxyribonucleotides [25].

1.2 RNA – Bridging the past and the present

1.2.1 The RNA World hypothesis

Despite attempts to discover exceptions to this, all known forms of cellular life utilize DNA to store their genetic information [26]. However, at a time predating cellular life, RNA presumably took on the role of the dominating genetic medium, coupling information and encoded function. This is known as the RNA world hypothesis [27–29]. Briefly, it links the emergence of life to the emergence of self-replicating RNAs, starting from single nucleotides and a sequential increase in sequence length and functional diversity of RNA (Figure 2A). While this process was non-enzymatic at first, driven by organic chemical reactions, the increase of functionality initiated the transition to ribozymes, catalytically active RNAs, as the crucial player [27,28]. Accompanying this process, the first proto cells entered the stage, eventually leading to the evolution of life as it is today [15,16]. Although this theory and several aspects of it are up to debate, recent studies contributed to a wide acceptance [29–31]. Thus, experimental data provides possible solutions to raised concerns including the prebiotically plausibility of the required chemistry, the limited catalytic repertoire and stability of RNA, compartmentalisation and how a transition to our current DNA/RNA/protein world might have happened.

The instability of RNA at elevated temperatures, basic pH or in presence of divalent metal cations, commonly found co-factors of ribozymes, has been considered prohibitive for the RNA world [29]. However, ribozyme activity was demonstrated in ice, at lower pH, and in absence of divalent ions, conditions where RNA stability is substantial and prebiotic RNA catalysis could have originated from [32–35].

Xu et al. demonstrated plausible pathways, leading to the formation of a four-membered potential alphabetic code, consisting of ribose- and deoxyribose-nucleosides [36]. Interestingly, this finding also provides an answer to how the complex stereochemistry, the three-dimensional orientation of atoms within a molecule, that we have today, arose from non-stereo-specific precursors [36,37]. There are numerous proposed routes for the subsequent phosphorylation, and chemical activation of these nucleosides, both prerequisites for the oligomerisation of nucleotides [38,39]. Finally, experimental data proved the ability of RNA for template directed polymerisation of oligomeric fragments or even monomers, as well as that these RNAs can emerge from a pool of random sequences [40,41]. From this RNA synthesis, the evolution of RNA-dependent RNA polymerases (RdRP) made up of RNA itself could have been achievable. Such self-replicating systems, able of open-ended evolution, potentially drove the transition from replicators depending on abiotic metabolic processes to autonomous systems with complex metabolic pathways [42,43].

The discovery of a ribozyme capable of catalysing the metabolically relevant aldol-reaction, and riboswitches utilizing diverse co-factors, highlight the essential chemical versatility of functions that primordial RNAs must have been able to perform in this case [44,45]. In addition, it could be demonstrated that the catalytically active part of ribozymes can be as short as five nucleotides. This drastically reduced the expected sequence space, required for the emergence of catalytic activity [46]. Although all these results suggest the validity of some sort of RNA world, it is impossible to exclude that RNA was not accompanied, coupled to, or even preceded by other possible chemicals capable of the same tasks [37,47]

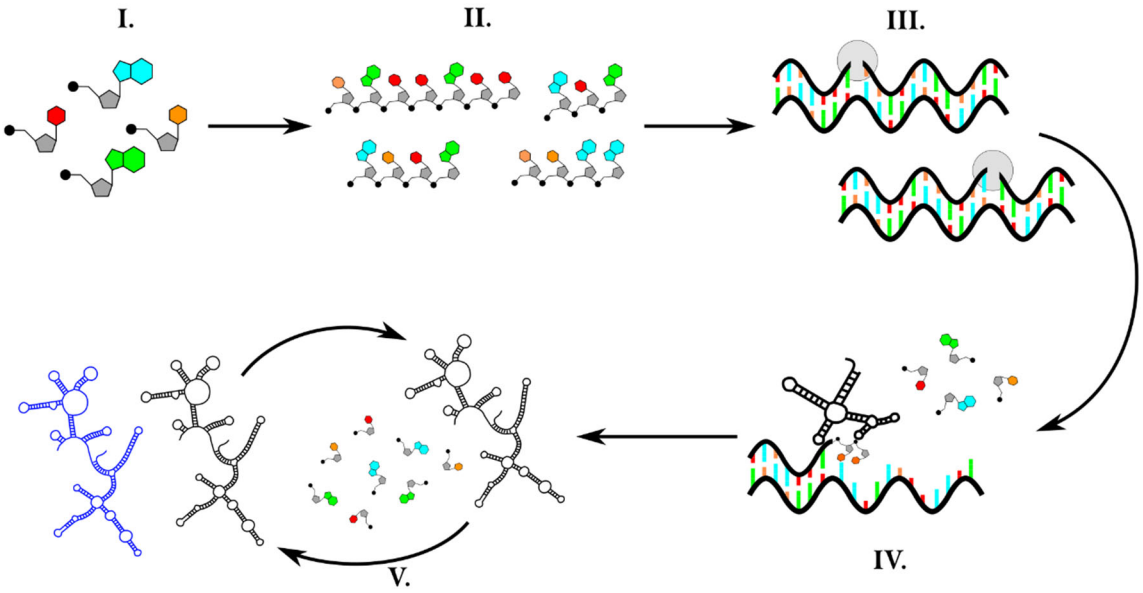


Figure 2 From small organic molecules to complex replicators: According to the RNA World theory, RNA played a crucial role during the early origin of life, potentially through the steps described in the following [27,28]. (**I.**) Ribonucleotides, consisting of ribose sugar (grey), a phosphate group (black) and one of several nucleobases (coloured), were generated by abiotic chemical synthesis. (**II.**) By some process, chemical activation of the phosphate groups occurred and allowed formation of nucleotide-oligomers. Ligation of short, activated oligomers then generated longer oligomers. (**III.**) At some point, longer strands began to facilitate the ligation of shorter ones, in a template directed manner. The template identity is established by sequence specific base-pairing with the smaller fragments and/or mononucleotides. This marks the begin of molecular evolution of RNAs, as sequences that facilitate ligation the most, produce more strands of identical sequence. In addition, the emergence of even longer strands is initiated (**IV.**) Continuing growth then allowed sequences to emerge, that allow RNA strands to adopt more diverse catalytically active structures. Thereby, some acquired the ability to not only elongate template fragments with short matching fragments but individual, activated nucleotides. These RNAs were the first RNA-dependent RNA polymerases. (**V.**) Finally, this led to the generation of polymerases that were capable to recognize and thus specifically replicate their own sequence, the first self-replicating RNAs. In the entailing evolution of these replicators, functionality, and genetic information increased, providing the respective entities with selection advantages. This eventually led to the emergence of primordial cellular life, ancestral to all present-day organisms [15]. In parallel, the RNA world made the transition to a DNA/protein/RNA world.

Noteworthy, proteins and ribozymes alike can employ co-factors to increase the range of chemical reactions they catalyse [45,48]. A prime example of a ribozyme utilizing co-factors, is the ribosome, a complex ribonucleoprotein particle (RNP) consisting of the ribosomal RNAs (rRNA) and a multitude of proteins. Here, the RNA performs the actual catalysis, the formation of peptide bonds during translation [49].The ribosome is often held as argument in favour of the RNA world, with the rRNA as a relic from the prebiotic era [50]. But it also serves as argument against it [29,30,51]. According to the 'protein first' hypothesis, the ribosome is the remnant of a protein dependent origin of life, where coded protein synthesis predates the emergence of RNA based replication. Phylogenetic analysis of the ribosomal proteins supposedly supports this assumption [30]. The ribosomal RNA would thus be a co-factor to the ribosomal proteins, and the ribosome would initially have acted as an RNA replicase [29].

However, an RNA world does not exclude that proteins synthesis and RNA replication evolved in parallel. Thus, the coevolution might actually have culminated in the coded protein synthesis using RNA to store the genetic information [50,52,53]. Interestingly, positively charged polypeptides have been shown to support ribozyme activity and locally enrich key components, while also providing a model for the emergence of protocells [54,55]. Interactions between negative charges in the RNA backbone and positive charges in the polypeptides can induce phase separation, effectively compartmentalising the RNA in droplets, the so-called coacervates [54,55]. Compartmentalisation is one of the most relevant steps towards cellular life, and crucial in preventing molecular parasites from leading to a population collapse [28,56]. Furthermore, the origin of the ribosome might be found in proto tRNAs, utilizing peptidic co-factors [50,52,53]. Hereby, peptide supported RNA replication/synthesis, and RNA catalysed peptide synthesis/aminoacylation might have resulted in a positive feedback loop. The respective interactions could have led to the emergence of longer chains of oppositely charged polymers, and in turn to a successive compartmentalisation in coacervates, amplifying this process even more. And while this is only a hypothesis, the idea of the ribosome as a 'frozen' coacervate, with the successors of the initial peptides kept in place by covalent links, is intriguing. The final step towards the current DNA/RNA/protein world, might then have been driven by the higher stability of DNA as genetic medium, required to adapt to a sudden increase in available genetic information. The transition from RNA to DNA could have been mediated by hybrid intermediaries, amplified through misincorporation of deoxyribonucleotides or reverse transcription [36,37].

1.2.2 RNA viruses

RdRPs belong to the oldest enzymes known, with a highly conserved catalytic centre [57,58]. *Farias et al.* place their origin at the dawn of protein translation, linking them to the replication of proto-tRNA genomes [53,57]. Furthermore, they suggest that the primordial RdRP was indeed

encoded in these proto-tRNAs [57]. Considering the age of this enzymatic group, they potentially even played a crucial role in the replication of putative RNA genomes of early cellular life [57,59]. Unfortunately, though, no fossils of these early cells have been found yet [60]. However, there is one group of biological entities, that still exists today and utilises RdRPs for the replication of their genomes – RNA viruses [57,58,61,62]. Many of them even carry tRNA-like structures within their genome, thereby resembling a form of proto-mRNA that is still capable of replication [53,63,64]. RNA viruses make up one of the most diverse group of biological entities [61]. The classification of viruses according to the Baltimore system distinguishes between seven groups, of which RNA viruses make up four groups – three that exclusively use RNA (group III, VI and V), one that also uses reverse transcribed DNA (group VI) [65]. Group III, IV and V differ in the architecture of their RNA genome, with group III having double stranded RNA (dsRNA), and group IV and V having single stranded RNA (ssRNA) genomes [65]. The difference between viruses belonging to group IV and V is the polarity of the genome. While members of group IV have a positive (+) polarity, that is their genomic RNA can directly serve as mRNA, group V is comprised from viruses with negative (-) polarity genomes, requiring the initial transcription of the complementary (+)-strand for protein translation [65].

Table 1: Overview of various types of RNA viruses.

Virus	Genome type	Genome size	Number of genes	Enveloped	Segmented
Emesvirus Zinderi [66]	(+)-ssRNA	3569 nt	4	no	no
Qubevirus [67]	(+)-ssRNA	4217 nt	4	no	no
SARS CoV-2 [68]	(+)-ssRNA	~ 29,900 nt	12	yes	no
Influenza A Virus [69]	(-)-ssRNA	~ 13,600 nt	8	yes	yes (8)
Ebola Virus [70]	(-)-ssRNA	~ 19,000 nt	7	yes	no
Rabies Virus [71]	(-)-ssRNA	~ 12,000 nt	5	no	no
Rota Virus A [72]	dsRNA	~ 18,500 bp	11	no	yes (11)
Bluetongue Virus [73,74]	dsRNA	~ 19,100 bp	10	no	yes (10)

The members of each group can be further distinguished, based on additional factors (Table 1), reflecting the sheer diversity within RNA viruses. Thus, they can package their genome in

protein shells of various shapes, called capsids, and some acquired an additional lipid layer envelop. Furthermore, the genome of some RNA viruses is segmented into multiple separate strands, similar to chromosomes, each encoding for its own specific set of genes. Although the encoded genes determine which organisms can be infected, high mutation rates enable fast adaption to novel hosts [75]. Consequentially, research interest in RNA viruses is high, as various members of them pose a high risk in form of human pathogenic agents. Some of the most devastating diseases are thus caused by RNA viruses, including Ebola, several forms of Hepatitis, Influenza, Corona virus disease 19 (Covid-19), measles and rabies [68,76,77].

1.2.3 Autonomy and self-replication

All viruses require at least the translational system of their host [78]. However, different viral species can exhibit varying degrees of dependence on host factors (Figure 3). Thus, viral autonomy inversely correlates with the dependence on host factors, especially regarding genome replication. Examples for low autonomy are viroids and satellite viruses. Viroids are small circular ssRNA pathogens of plants, with genomes in the range of 200 to 400 nucleotides (nt). Although they do not encode for any peptides, they can have devastating effects on infected hosts [79,80]. And as they do not encode for a peptide, they completely depend on host factors for the replication and spread of the RNA genome. Satellite viruses have genomes with sizes of up to 1500 nt length, depending on the the virus type.

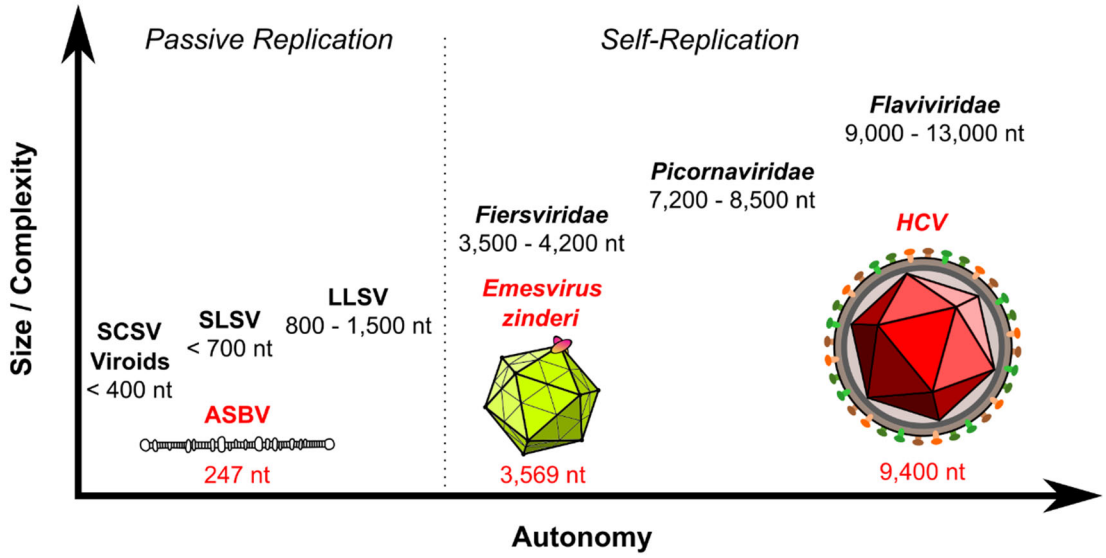


Figure 3 Correlation between size/complexity and autonomy: Examples of replicating RNA species, separated into passive and self-replication. Here, passive replication is defined as the absence of genes encoding for the catalytically active components of the replication machinery. In contrast, self-replication is defined as the presence of genes encoding for the main catalytic replication component, that is an RdRP. Passive RNA replicators include viroids, small circular (SCSV), small linear (SLSV) and large, linear satellite viruses (LLSV). Avocado sunblotch viroid (ASBV) is given as an example for viroid RNA [81], *Emesvirus zinderi* for non-enveloped RNA bacteriophages of the *Fiersviridae* family [66], and Hepatitis C virus (HCV) for human pathogenic, enveloped RNA viruses [82].

While small circular satellite viruses (SCSV) do not, small, and large linear satellite viruses (SLSV / LLSV) can encode for peptides, involved in replication and encapsidation [83]. However, so far none of the identified genes encodes for an RdRP [84]. Differently to viroids, satellite viruses do not depend on host factors alone. Instead, they require the replication machinery of their respective helper virus, co-infecting the same host cell [83]. In general, both viroids and satellite viruses can be considered passive replicators, RNA species that do not encode the function for their own replication.

In contrast, larger RNA viruses usually encode at least the catalytically active replicase subunit in their genome. RNA bacteriophages of the *Fiersviridae* family, with genomes only three times larger than LLSVs, already encode up to four genes providing a substantial degree of autonomy [85]. These genes include proteins for capsid formation, host infection and lysis, and most importantly an RdRP subunit [66,86,87]. Consequentially, these viruses can already be considered self-replicating RNA genomes. Interestingly, they are among, or even the most ancient viruses [58]. Thus, their RdRPs exhibit high structural homology to the putative primordial progenitor described by *de Farias et al.* [57]. The members of the *Flaviviridae* family, including Hepatitis C virus, are up to three times larger than *Fiersviridae* [82,88]. Compared to the smaller RNA phages, their genome encodes a fully active RdRP, and additional factors to actively recruit host factors to so-called replication organelles [89,90].

Noteworthy, the correlation between genome size and increased functionality/autonomy eventually weakens, once a certain threshold was passed. In detail, a coding sequence of a certain length can encode only for a limited number of amino acids. This lies rooted in how the genetic code is translated into an amino acid sequence. Thereby, three nucleotides correspond to one amino acid [91,92]. In addition, the reading frame of these triplets can be shifted, by +1 and +2 nucleotides. Then it reverts to the original reading frame, shifted by one triplet. Thus, if the full potential of frame overlap is used, a sequence of N nucleotides can encode for maximal N-2 amino acids (N/3 per reading frame, minus 1 per shifted frame). However, especially in larger viral genomes, this tends to be not the case, effectively reducing the information density and potential autonomy [62]. Furthermore, an increase in genome size is accompanied by an increased need for factors to maintain the genome integrity. This is due to the in general high error rates of RNA replicases [62,93]. Longer genomes increase the chances that lethal mutations occur. In combination with the already mentioned instability of RNA under physiological conditions [29], it was speculated that the size of RNA genomes is in fact limited [93,94].

Nonetheless, the self-replication of RNA genomes poses an interesting subject for research. Following the global outbreak of the Covid-19 pandemic, the first mRNA-based vaccines were approved for their application in humans [95]. But even before this, mRNAs were investigated as basis for potential vaccine development [95]. Thereby, self-replication of vaccine-mRNAs is currently tested for its potential to boost antigen expression and thus the immune response

[96,97]. Furthermore, *Yoshioka et al.* successfully used self-replicating RNAs to reprogram human cells, generating induced pluripotent stem cells [98]. Finally, as aforementioned, RNA replisomes might have played a crucial role in the later stage of the origin of cellular life [53,57]. Consequentially, synthetic RNA replisomes might shed light on how this role was filled out and how evolution of increasingly large genomes might have taken place. Similarly, RNA based genomes even facilitate approaches to design minimal cells in synthetic biology, reducing the number of biological macromolecule species from three (DNA/RNA/protein) down to two (RNA/protein) [85,99]. Especially viruses of the *Fiersviridae* family could excel here, due to their small genome size and dependence on translation factors, that are already present in *in vitro* translation systems [66,100–102].

1.3 *Emesvirus zinderi* – What is known and what not?

1.3.1 Life cycle of a ssRNA virus

Emesvirus zinderi, in the following referred to as bacteriophage MS2, is member of a group of small, single stranded RNA (ssRNA) viruses, belonging to the family of *Fiersviridae* [103].The genome of these viruses has positive polarity, it thus can directly serve as mRNA for the expression of the viral proteins [103]. Noteworthy, the MS2 genome is one of the smallest known genomes, with a size of 3569 nucleotides, encoding for only four genes (Figure 4) [66,104]. Two of the four gene products, the maturation protein (MP) and the coat protein (CP), form the viral capsid, protecting the viral RNA and mediating the infection of host cells [105]. Thereby, 89 CP dimers and one MP assemble into the icosahedral shell of a single RNA [105,106].

In the first step of host infection, the capsid binds via the MP to the F-pili of male *Escherichia coli* [105,107]. These pili, used for the transfer of genetic elements between bacteria through conjugation, can also serve as entry point for various bacteriophages [107–111]. In general, F-pili are dynamic structures, connecting the cytoplasm with the extracellular space, thereby crossing both cell membranes and the periplasm [107,110]. Usually, the extension and retraction, respectively, serve to find and subsequently draw in mating partners, followed by transfer of genetic material like the fertility factor into the recipient cell [108,110,112]. However, phages like MS2 exploit this system for infection. While the mechanism is not yet fully understood, RNA import presumably occurs by retractile pili pulling the MP/RNA-complex out of the capsid and across the membranes into the cytoplasm [113]. After the import of genomic MS2 RNA and dissociation of the MP, translation of the CP and RNA dependent replicase subunit (MS2rep) from the respective genes (*cp* / *rep*) soon takes place [114–116]. This is then quickly followed by replication of the RNA genome, as well as translation of both the MP and the lysis protein (LP) [66,115]. Thereby, MS2rep forms the active complex with host factors, including elongation factors Tu (EF-Tu), Ts (EF-Ts), ribosomal protein S1 and a putative, unknown factor [66,117].

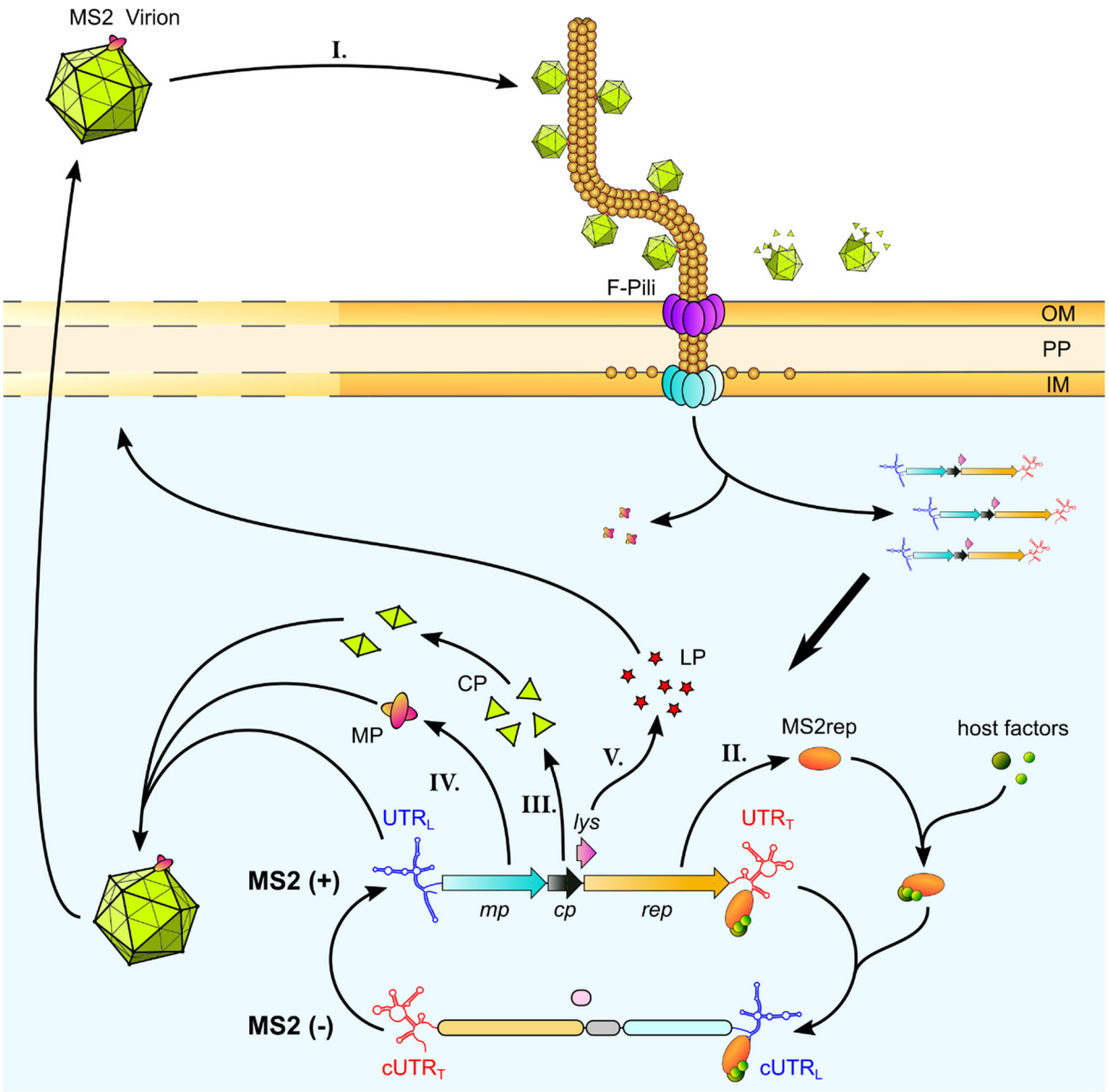


Figure 4 Life cycle of bacteriophage MS2: (**I.**) The MS2 genome is packed into capsids. These bind to the F-Pili of *E. coli* host cells. Retraction of these pili into the cytosol, leads to the cellular import of genomic MS2 RNA, still bound by the maturation protein. Thereby the RNP crosses the outer cellular membrane (OM), the periplasm (PP) and finally the inner membrane (IM). In this process, shedding of the empty coat protein shell occurs. At some point, the MP-RNA complex dissociates. The genomic RNA serves as mRNA for the expression of the four MS2 genes (arrows), flanked by the leading 5'-UTR (UTR$_L$) and the tailing 3'-UTR (UTR$_T$). (**II.**) Translation of the *replicase* gene (*rep*) generates the replicase subunit, which subsequently binds *E. coli* host factors. The active replicase complex uses the genomic strand (MS2(+)) as template for the synthesis of the antigenomic strand (MS2(-)) and vice versa. Thus, the MS2 genome gets amplified. (**III.**) Expression of the *coat protein* gene (*cp*) produces monomeric coat protein (CP). Following dimerization, CP-dimers bind to regulatory and packaging motifs in the genomic MS2 RNA. (**IV.**) Replication-linked restructuring of RNA motifs allows ribosome binding to the ribosome binding site upstream of the *maturation protein* gene (*mp*). Binding of the maturation protein (MP) is accompanied by cooperative condensation of the RNA genome, bound by CP dimers. Consequentially, the RNA gets packaged into the mature capsid. (**V.**) Translation from the *cp* gene allows in parallel expression of the *lysis* gene (*lys*). This eventually leads to cell lysis and finally to the release of mature MS2 virions.

Starting from the genomic (+)-strand, the active replicase first synthesises the intermediary antigenomic (-)-strand, before synthesising a new (+)-strand. Expression of the MP leads to formation of mature MS2 virions [105,106,118]. And finally, cell lysis, mediated by the LP, enables release of these virions, and allows for infection of new host cells. The underlying mechanism of this lysis event however is not yet fully clear. So far, it only was possible to deduce that it depends on the chaperone and protein-disulfide isomerase DnaJ and presumably interferes with cell wall synthesis. [119–122]

The genome of MS2, just based on its size and information encoded by RNA sequence, might be considered small and simple. However, small ssRNA viruses are notorious for genome compaction, resulting in abundance of overlapping open reading frames (ORF), as well as high information content of structural RNA elements [62,123]. Thus, the *rep* gene of MS2 is regulated by the interaction of a repressor hairpin in the RNA with a CP dimer. Upon binding, this interaction disables expression of more MS2rep [124]. In addition, the expression of the rep gene is coupled to the translation of the CP [125]. The expression of the *mp* gene is only possible during a short time window after synthesis of a nascent (+)-strand. Only then, when secondary structures have not yet fully formed, the ribosome binding site (RBS) upstream of the *mp* coding region is accessible and translation can be initiated [126]. Similarly, expression of the CP is regulated by RNA motifs upstream of its coding region [127]. In case of the *lys* gene, it partially overlaps with the ORFs of the *cp* and the *rep* genes, albeit its ORF is shifted by +1 base pair (bp) relatively to the *cp* [128,129]. The translation from the *lys* ORF is limited by its dependence on preceding translation of the coat protein. Hereby, translation of the coat protein disrupts the structure of an inhibitory RNA hairpin [128]. It is also essential, as translation of the lysis protein is coupled to a frameshift and premature translation termination of ribosomes translating the *cp* ORF [129]. Finally, even the packaging of the genomic RNA, and likely also RNA replication, depend on distinct RNA hairpins. In case of packaging, binding of the MP and multiple CP dimers requires certain RNA hairpins, and only the cumulative effect of these individual binding events allows for the condensation of the RNA into the capsid lumen [130,131].

1.3.2 MS2 – Small virus, big potential

MS2 is one of the best studied model organisms in modern molecular biology. The MS2 coat protein was the first fully sequenced gene [132]. In 1976, *Fiers et al.* [66] then published the last missing sequence of the MS2 genome, thus it was the first genome to be fully sequenced, furthermore the first RNA genome as well. The first DNA genome, the 5375-nucleotide long genome of bacteriophage φX174, was fully sequenced only one year later, in 1977 [133]. These events allowed for the systematic search for coding sequences and started an era of genome sequencing and analysis, currently peaking in the identification of the last missing 8% of the 3 billion bp large human genome [134].

Furthermore, since its first discovery in 1961 [135], MS2 as a model system contributed to the unravelling of numerous biological processes. Thus, *Kastelein et al.* could demonstrate the effect of 5'- and 3'- untranslated regions (UTR) on the translation efficiency, using the coat protein cistron as a model [136]. In a more general sense, as a readily available model for mRNA, the MS2 genome helped to understand further important processes underlying protein translation, ranging from the role of secondary structures, RNA folding and long-range interactions on translation efficiency, to general translation initiation [124,126,127,137,138]. In addition, MS2 is a ssRNA virus that can be handled without costly equipment or sophisticated precautions. Hence, MS2 also provided insights into the life cycle, and enabled the development of tools for studies, respectively, of human pathogenic viruses [139–143] Especially the coat protein has found use in biotechnology. Exploiting the specific interaction of the translational repressor hairpin of the *rep* gene with the CP-dimer, it was possible to detect and localise RNAs *in vivo* [144]. Besides that, the CP is also currently investigated for its capsid-forming potential *in vitro*, for example to generate armored RNA particles [145,146], drug delivery [147–149], or the development of novel vaccines [150–152].

However, even though MS2 has been studied so well, there are numerous aspects of it that have escaped further elucidation. And with three more proteins next to the coat protein, its potential for research and biotechnological/medical applications has surely not been exhausted yet.

1.4 Of Parasites and Mutualists

The question if viruses are alive has been disputed since the early 20th century [153,154]. Replication and evolution of viruses are held as arguments in favour of this question, while their lack of an own metabolism is often regarded as a criterion against it [153]. Although the International Committee on Taxonomy of Viruses has officially acknowledged, that viruses are not alive [155] in 1999, recent discoveries keep the debate ongoing. Based on comparative structural and functional analysis, *Nasir et al.* suggest that viruses evolved by reduction from so-called proto-virocells, and potentially constitute their own branch in the tree of life, a fourth domain next to archaea, bacteria and eukarya [11]. This is supported by the discovery of giant viruses, with genomes even exceeding the size of some bacterial genomes, while also encoding for a set of metabolic genes [156–158]. In contrast to both stances, *Koonin* and *Starokadomskyy* dismiss the question in general, as misguided, and non-scientific – the answer always depending on the definition of life [159]. Furthermore, considering the tremendous diversity among viruses, a generalised classification appears impractical at best and impossible at worst [160,161].

Nonetheless, that viruses are usually parasitic entities, with a substantial effect on the whole biosphere, was in general agreed upon [11,56,154,159,162]. More precisely, viruses are obligate

intracellular parasites. That means that they depend on penetration into the host organism's cytoplasm. Only there, all the conditions are met that allow for the completion of the viral life cycle, genome replication and encapsidation [163]. Thus, viruses are considered inactive, once they are outside their host cell, though some exceptions exist [164]. Inside their target cell, viruses hijack the host's translational system to replicate their genome and propagate [78,165]. This is usually followed by the phage escape, sometimes mediated by the lysis and death of the host cell [166,167]. Hence, viruses can be best described as genetic/molecular parasites, bound to their host in a consumer-resource relation, where the consumer (virus) draws resources (metabolites, enzymes) from the host [56,162,168]. This is highly detrimental for the host, and thereby functions as negative selection pressure. Consequentially, it acts as an evolutionary driver for both the host and the virus, trapping them in a constant arms race [169–171].

But parasitism is only one form of how organisms can interact. The opposite of parasitism is mutualism. While in a parasitic relation, one partner thrives at the cost of the other, a mutualistic relation benefits both partners [172]. As stated above, viruses are in general considered to be parasitic. However, the sheer variety of viruses allows for exceptions, and eventually the view on them as 'antagonists only' must be updated, as more and more mutualistic virus-host interactions are discovered. By enabling the host to open up new resources or providing more cost-efficient means to access and manage existing ones, respectively, both partners can gain a substantial selection advantage [168]. Thus, bacteria can deploy phage infection as regulator of population-size, countermeasure against competitors and a source of new genetic material through HGT [173]. But eukaryotic hosts have been shown to benefit from interactions with certain viruses as well [174,175]. The evolution of the placenta in mammals, for example, was mediated by the endogenization of mutualistic retroviral elements into their genome [176]. The acquisition of the placenta then presumably provided a selection advantage by increasing foetal and maternal fitness [177]. In other cases, viruses can contribute to the immune defence of their host or provide tolerance to environmental stress [174]. All in all, however, these mutualistic host-virus relations depend on a well-regulated viral life cycle [178]. Disturbance of this might easily lead to a shift from mutualism to parasitism [175,178,179].

1.5 Reprogramming MS2 – Why and How?

The bacteriophage MS2 is an RNA parasite, infection of it is detrimental to the host. The goal of this study is the attempt to reprogram MS2 from a harmful parasite to a beneficial mutualist. The simplest form of a mutualistic RNA species could already be constituted by a self-replicating RNA, encoding for the RdRP subunit and a short selection marker like the resistance gene against the antibiotic Zeocin. But furthermore, this study aims at evaluating the potential of MS2 for the design of self-replicating RNAs in general and, if possible, synthetic RNA genomes. This will contribute to answering questions regarding the following topics:

1. **Potential of RNA as genetic medium during the origin of life**
2. **Evolution of RNA viruses and their impact on host evolution**
3. **RNA genomes for minimal cells in synthetic biology**
4. **Infection pathways of non-enveloped viruses**
5. **Applications in biotechnology**

In detail, the hypothetical RNA genome is based on two genes of the MS2 genome, each an integral part of its respective module. The first module, for RNA delivery is based on the MP and its binding RNA motif. Second, the replicase subunit, to enable replication of the RNA genome. Thereby, the path towards each module will be based on current knowledge and methods presented below.

Previous work of *Weise et al.* demonstrated activity of *in vitro* translated MS2 replicase subunit in the translational system PURExpress (NEB) [180]. PURExpress is an adapted version of the original PURE (protein synthesis using recombinant elements), described by *Shimizu et al.* [181]. In short, the PURE system contains all factors required for *in vitro* translation (30 translation factors / ribosomes / 46 tRNAs), *in vitro* transcription (IVT) by T7 and a system for NTP regeneration using creatine phosphate [181]. Thereby, each protein component is individually purified, thus making the PURE system a well-defined environment. The well-defined nature of these systems makes them excellent starting points for the reconstitution of enzymatic activity *in vitro* and identification of co-factors, in this case of the MS2 replicase. Consequentially, as activity of MS2rep was demonstrated in PURExpress, it must contain all essential co-factors.

Furthermore, *Weise et al.* established a read-out for this activity, which will serve as the foundation for initial activity verification [180]. This read-out is based on fluorescence changes that are coupled to the transcription of new RNA from the non-fluorescent input RNA, [F30-Bro(-)]$_{UTRs(-)}$. Hereby, the fluorescence output is generated from the broccoli aptamer. The broccoli aptamer, which is part of the newly synthesised RNA, binds to the fluorogenic dye DFHBI-1T ((5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)-methylene]-3,5-dihydro-2-methyl-3-(2,2,2-trifluoroethyl)-4H-imidazol-4-one). When excited with blue light, the bound DFHBI-1T exhibits green fluorescence [182,183]. The reconstitution of the MS2 replicase complex allows for the identification of replication scaffolds, and subsequently RNA genomes.

In case of the maturation protein, previous reports on the structure of the MS2 capsid have shed light on the crucial interactions and components for the infection of *E. coli* by MS2 [105,107]. Furthermore, it was shown by *Katanaev et al.*, that an RNP consisting of only the MP and genomic MS2 RNA, so-called minimal infectious units, are already sufficient to mediate substantial infection of host cells [184]. The MP, as a structural protein, does not have any known enzymatic activity. But monitoring the infection of host cells with *in vitro* assembled minimal infectious units allows to verify the functional integrity of recombinantly purified MP. The

purification of functional MP is thereby the first step to establish the delivery module.
Once this step is completed, binding RNA motifs could be identified by systematic evolution of ligands by exponential enrichment (SELEX) [185–187]. In the first step of SELEX, an RNA pool of randomized sequences is incubated with an immobilized target. A fraction of these sequences will allow the RNAs to adopt secondary structures that can bind to the target, while RNAs that do not or only transiently bind, can be removed [187]. Reverse transcription and a following polymerase chain reaction (PCR) of the bound sequences lead to an exponential amplification of them. A subsequent IVT step generates a new pool of RNAs for the next iteration of the cycle. With each following iteration, the RNAs that bind the strongest will sequentially enrich, outcompeting those that do not.

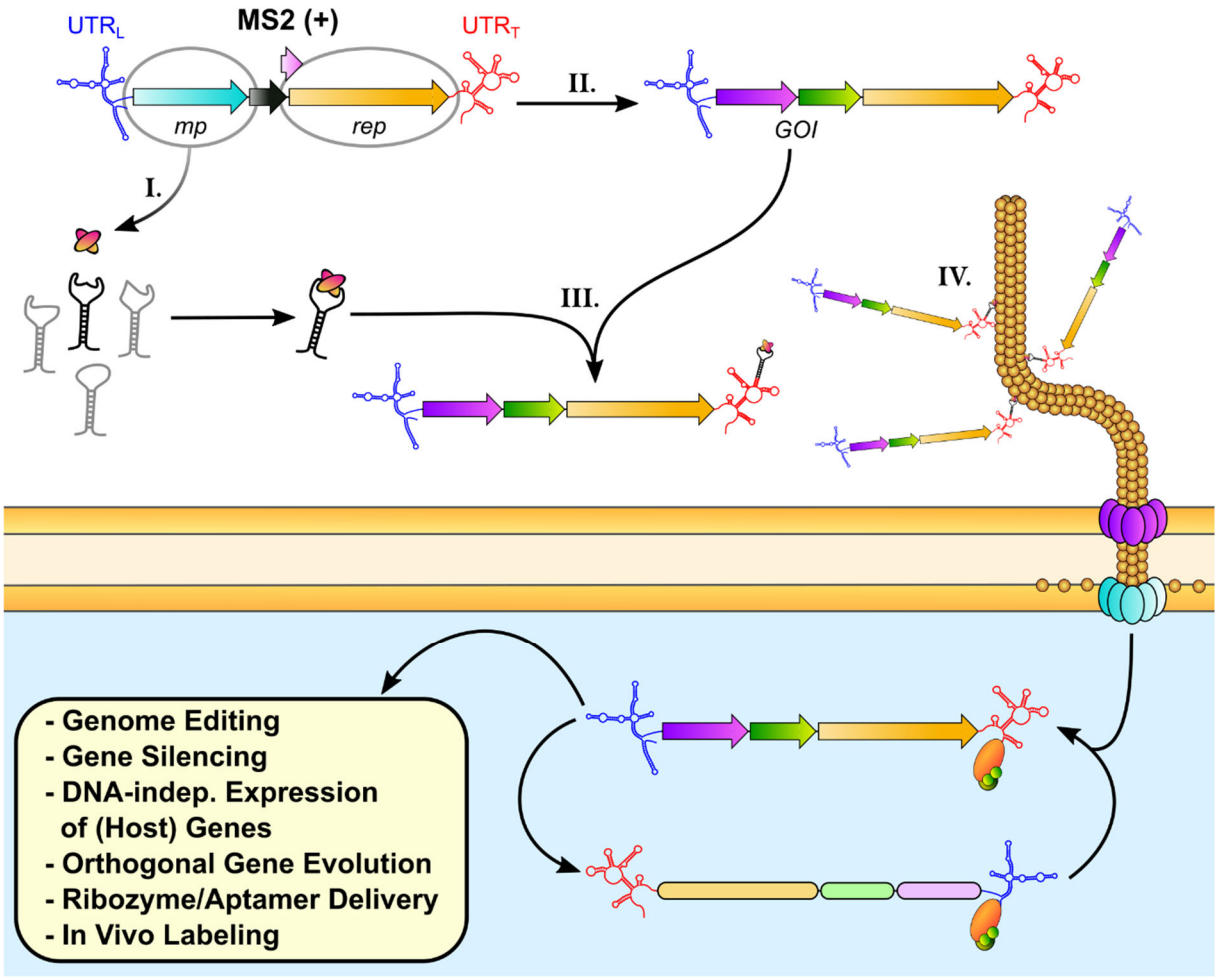


Figure 5 Design of a self-replicating and self-transforming RNA genome: Using the infection and replication machinery of MS2, an RNA genome is sought to be generated. This genome can ideally replicate *in vivo* and be transformed into MS2-susceptible *E. coli*. (**I.**) In a first step, the MP gets purified and used for the selection of MP-binding RNAs. (**II.**) In parallel, the MS2 RNA can be used as a scaffold to build a synthetic RNA replicator, encoding for the *rep* gene and additional genes of interest (*GOI*) (**III.**) The final RNA genome is then assembled from the replicator scaffold, the MP, and the MP-binding motifs. (**IV.**) Finally, using the F-pili of male *E. coli*, the pre-assembled RNP can be delivered into cells. There it gets replicated and expression of the encoded *GOI*s can perform the desired function.

In the past, this approach has been successfully used to generate nucleic acid aptamers binding to enzymes, amino acids, or organic dyes [183,186,187].
Combining the two modules, the first synthetic RNA replisome can be assembled (Figure 5). Thereby, a variety of functions can be encoded in the replisome. These functions can range from enzymes and components for gene engineering and RNA interference, or simply for the overexpression of exogenous genes. Potentially, even ribozymes and/or aptamers can be encoded, providing even further biological activity. Interestingly, it could also serve as self-replicating mRNA for the expression of genes exogenised from the host genome. Thereby, the synthetic, self-amplifying RNA would act as a genomic transplant. Together with the high mutation rates inherent to viral RdRPs, new approaches to directed *in vivo* evolution are possible. Nonetheless, this is only the first step in using MS2 to address the topics mentioned before.

In summary, the main goals of this work were the following:

- **Design a MP-based system for the delivery of synthetic RNAs**
- **Reconstitute and characterise the MS2 replication complex *in vitro***
- **Identify replicating RNA scaffolds**
- **Combine the delivery and replication module into a self-replicating RNA**
- **Expand the encoded genetic information of self-replicating RNAs**
- **Evaluate the potential of *in vivo* replication of synthetic RNA genomes**

2. Materials and methods

2.1 Materials

2.1.1 Chemicals and suppliers

Table 2 shows all chemicals used for this thesis, as well as their common abbreviations and their supplier. If not indicated differently, percentages always refer to volume per volume for liquids and weight per volume for solid chemicals.

Table 2: Chemicals and suppliers

Chemical (Abbreviation)	Supplier
3-(N-morpholino)propanesulfonic Acid (MOPS)	Roth
4-(2-Hydroxyethyl)-1-piperazineethanesulfonic Acid (HEPES)	Roth
5-Brom-4-chlor-3-indoxyl-β-D-galactopyranosid (X-Gal)	Roth
(5Z)-5-[(3,5-Difluoro-4-hydroxyphenyl)methylene]-3,5-dihydro-2-methyl-3-(2,2,2-trifluoroethyl)-4H-imidazol-4-one (DFHBI-1T)	Jena Bioscience
Acetic Acid (HOAc)	Sigma Aldrich
Adenosine triphosphate (ATP)	Thermo Scientific
Ammonium Chloride	Roth
Ammonium Persulfate (APS)	Roth
Boric Acid	ChemSolute
Bromophenol Blue	Roth
Carbenicillin Sodium Salt	Serva
Chloramphenicol	Sigma Aldrich
Coomassie Brilliant Blue R250	Roth
Cytidine triphosphate (CTP)	Thermo Scientific
D-(+)-Glucose	Sigma Aldrich
D-(+)-Maltose Monohydrate	Fisher Bioreagents
Dimethyl Sulfoxide (DMSO)	Serva
Dithiotreitol (DTT)	Fisher Bioreagents
Ethanol	ChemSolute
Ethylenediaminetetraacetic Acid (EDTA)	Roth
Formamide	Roth
Glycerol	Roth
Glycin	Sigma Aldrich
Guanosine-5'-triphosphate (GTP)	Thermo Scientific
Hydrochloric Acid	VWR Chemicals

Imidazole	Roth
Isopropyl β-D-1-Thiogalactopyranoside (IPTG)	Fisher Bioreagents
Kanamycin Sulfate	Roth
L-(+)-Arabinose	Roth
Lactose Monohydrate	Merck
Magnesium Chloride	Merck
Manganese Chloride Monohydrate	Roth
Nicotinamide Dinucleotide (NAD^+)	New England Biolabs
Phenylmethylsulfonyl Fluoride (PMSF)	PanReac AppliChem /ITW Reagents
Polyacrylic Acid Sodium Salt (MW ~2.100)	Sigma Aldrich
Polyethyleneglycol MW ~8000	Roth
Potassium Chloride	Roth
Potassium Hydroxide	Sigma Aldrich
Rifampicin	Roth
Rubidium Chloride	Roth
Sodium Chloride (NaCl)	Merck
Sodium Dodecyl Sulfate (SDS)	Sigma Aldrich
Sodium Hydroxide	VWR Chemicals
Streptomycin Sulfate	Corning
SYBR Gold	Invitrogen
SYBR Safe	Invitrogen
Tetracycline Hydrochloride	Sigma Aldrich
Tetramethylethylenediamine (TEMED)	Roth
Thymol Blue	Roth
Tricine Pufferan	Roth
Tris Hydrochloride	PanReac AppliChem /ITW Reagents
Tris(hydroxymethyl)aminomethane (Tris)	Th. Geyer
Triton X-100	Sigma Aldrich
Tween 20	MP Biomedicals
Urea	Roth
Uridine triphosphate (UTP)	Thermo Scientific
Xylene Cyanol FF	Sigma Aldrich
Zeocin	Invitrogen
β-Mercaptoethanol (β-ME)	Merck

2.1.2 E. coli strains

The *E. coli* strain Top10 (Invitrogen) was used for propagation of plasmids, as well as for the AraC regulated overexpression of proteins from inducible pBAD33 expression vectors.
For protein overexpression from inducible pET28a plasmids, under control of LacI and T7 RNA polymerase, the *E. coli* strain BL21(DE3) (Thermo Scientific) was used, which was kindly provided by the core facility at the Max-Planck Institute of Biochemistry (Martinsried, Germany). Additionally, the BL21 (DE3) strain was also used for overexpression of proteins from pLD1, pLD2 and pLD3.[188]
All experiments involving delivery of *Emesvirus zinderi* RNA were performed using the *E. coli* strain F^+5695 (ATCC 12453) or self-made derivatives of this strain. The F^+5695 derived strains and the individual preparation of each of them are listed and described in detail in the respective sections.
The *E. coli* HMS-174(DE3) strain was purchased from Merck. The original strain and all derivatives of it were used for *in vivo* studies of the EcI5 intron. Genomic integration of *sacB* gene into the *rnc* locus is described in section 2.5.

2.1.3 Media and antibiotics

In general, all steps that required bacterial growth media, both liquid culture and plates, were done using Lysogeny Broth (LB Lennox). The medium was prepared by dissolving 10 g Tryptone (Roth), 5 g Yeast Extract (Roth) and 5 g NaCl per liter water and sterilized by autoclaving. Culture plates were prepared using BD Difco Bacto Agar (Fisher Scientifc) in a final concentration of 1.5 wt.-%. Table 3 lists all used antibiotics, as well as their standard working concentrations and which plasmid or, respectively strain, they were used for.

Table 3: Antibiotics, applications, and concentrations.

Antibiotic	**Plasmid**	***E. coli* Strain**	**Concentration [µg/mL]**
Carbenicillin (Carb)	pEX-A124, pGEM-T Easy, pMini, pUCIDT,	-	100
Chloramphenicol (Cam)	pBAD33	-	34
Kanamycin (Kann)	pET28a, pLD1, pLD2, pLD3	HMS-174(DE3)^{sacB}	50
Streptomycin (Strep)	-	F^+5696	2
Rifampicin (Rif)	-	HMS-174(DE3)	100
Zeocin (Zeo)	-	-	37.5

2.1.4 Buffers

All buffers for protein purification and assays designed for this study are listed with their composition in Table 4, and sorted by application.

Table 4: Buffer compositions and applications.

Purification of Maturation Protein	
LyB1	HEPES pH 8.0 (50 mM), NaCl (500 mM), EDTA pH 8.0 (1 mM), DTT (1 mM), PMSF (1 mM), Sodium Polyacrylate (1g/100 mL), Glycerol (5 %)
WBL1	HEPES pH 8.0 (50 mM), NaCl (500 mM), EDTA pH 8.0 (1 mM), DTT (1 mM), Sodium Polyacrylate (1g/100 mL), Glycerol (5 %)
EB1	HEPES pH 8.0 (50 mM), NaCl (500 mM), EDTA pH 8.0 (1 mM), DTT (50 mM), Sodium Polyacrylate (1g/100 mL), Glycerol (5 %)
DB1	HEPES pH 8.0 (50 mM), NaCl (500 mM), EDTA pH 8.0 (1 mM), DTT (5 mM), Sodium Polyacrylate (1g/100 mL), Glycerol (25 %)
Purification of MS2rep, Qβrep and co-factors	
LyB2	HEPES pH 7.5 (50 mM), NH_4Cl (250 mM), $MgCl_2$ (10 mM), DTT (1 mM), PMSF (1 mM), Imidazol (20 mM)
WBL2	HEPES pH 7.5 (50 mM), NH_4Cl (250 mM), $MgCl_2$ (10 mM), DTT (1 mM), Imidazol (20 mM)
WBH2	HEPES pH 7.5 (50 mM), NH_4Cl (1000 mM), $MgCl_2$ (10 mM), DTT (1 mM), Imidazol (20 mM)
EB2	HEPES pH 7.5 (50 mM), NH_4Cl (250 mM), $MgCl_2$ (10 mM), DTT (1 mM), Imidazol (300 mM)
DB2	HEPES pH 7.5 (50 mM), KCl (100 mM), $MgCl_2$ (10 mM), DTT (5 mM), Glycerol (30 %)
Purification of hyperactive MuA transposase	
LyB3	HEPES pH 8.0 (50 mM), NaCl (500 mM), EDTA pH 8.0 (1 mM), DTT (1 mM), PMSF (1 mM)
WBL3	HEPES pH 8.0 (50 mM), NaCl (500 mM), EDTA pH 8.0 (1 mM), DTT (1 mM)
WBH3	HEPES pH 8.0 (50 mM), NaCl (1250 mM), EDTA pH 8.0 (1 mM), DTT (1 mM)
EB3	HEPES pH 8.0 (50 mM), NaCl (500 mM), EDTA pH 8.0 (1 mM), DTT (50 mM), Glycerol (10 %)
DB3	HEPES pH 7.5 (50 mM), NaCl (500 mM), EDTA pH 8.0 (1 mM), DTT (5 mM), Glycerol (25 %)
Buffers for *in vitro* & *in vivo* Assays	
5x PBB	HEPES pH 8.0 (100 mM), $MgOAc_2$ (50 mM)
6.7x EMSA B1	Glycerol (80 %), SYBR Gold (3x), Thymol Blue (0.03 %)
2x EMSA B2	Glycerol (50 %), SYBR Gold (2x), Thymol Blue (0.01 %)
1x RIMB	Tric·HCl pH 8.0 (25 mM), Glycerol (15 %), Triton X-100 (0.05 %), NaCl (126 mM), $MgCl_2$ (10 mM)
Buffers for Electrophoresis	
50x TAE	Tris (2 M), HOAc (1 M), EDTA (50 mM)
10x TA	Tris (400 mM), HOAc (200 mM)
10x TBE	Tris (890 mM), Boric Acid (890 mM), EDTA (20 mM)
1x TBU-X	Tris (89 mM), Boric Acid (89 mM), EDTA (2 mM), Urea (7 M), Acrylamide/Bisacrylamide 19:1 (X %)
10x TB_{acidic}	Tris·HCl (0.5 M), Boric Acid (0.5 M)
ResB-X	Tris·HCl pH 8.8 (375 mM), SDS (0.1 %), Acrylamide/Bisacrylamide 37.5:1 (X %)

StaB	Tris·HCl pH 6.8 (125 mM), SDS (0.1 %), Acrylamide/Bisacrylamide 37.5:1 (5 %)
10x SDS-RB	Tris (250 mM), Glycine (2 M), SDS (1 %)
2x SDS Loading Buffer	SDS (4 %), Glycerol (20 %), DTT (200 mM), Tris·HCl pH 6.8 (125 mM), Bromphenol blue (0.004 %).
Additional Buffers	
TB1	$RbCl_2$ (100 mM), $MnCl_2$ (50 mM), KOAc (30 mM), $CaCl_2$ (10 mM), Glycerol (15 %), adjusted to pH 5.8 with HOAc
TB2	$RbCl_2$ (10 mM), $CaCl_2$ (75 mM), Glycerol (15 %), MOPS (10 mM), adjusted to pH 6.5 with KOH

2.2 Devices and software

For imaging of plates from plaque assays and gels from electrophoresis of proteins, a Moluecular Imager Gel Doc XR+ system (Biorad) was used, and the raw data was analyzed with Image Lab software v.5.2.1 (Biorad).

In the case of evaluating the inhibition effect of IF3 on *in vitro* RNA replication by MS2rep, time resolved fluorescence assays were performed in a T-16 ISO fluorescence system (AXXIN) using its desktop application version 1.8.1.3. All other time resolved fluorescence assays were performed in a StepOne Real-Time PCR System (Thermo Fisher Scientific) with the StepOne and StepOnePlus Software v.2.3.

The fluorescence anisotropy experiment was performed in a CLARIOstar Plus micro plate reader (BMG Lab Tech), and the data subsequently analysed using MARS software v.4.01 R4 (BMG Lab Tech).

All gels from electrophoresis of nucleic acids were imaged using a Microtek Bio-1000F scanner with Microtek MiBio Fluo v.1.04 software (Microtek).

Lysis of cells for protein purification was achieved by sonication in ten minutes increments using a Sonopuls sonicator (Bandelin) with a TS 106 tip (67 % amplitude, 15 seconds on /15 seconds off).

Processing of raw data from deep sequencing of RNA hairpins was done with Cutadapt v1.9, using a quality threshold for the base reads of Q30 (>99.9% fidelity).

The data and figures presented in this thesis were analysed, visualized, and prepared using GraphPad Prism v.9.2.0, as well as Inkscape v.1.0.1.

2.3 General methods

2.3.1 Gel electrophoresis of nucleic acids

For general agarose gel electrophoresis of DNA and RNA, the gels were prepared using 1x TAE buffer and Agarose Standard (Roth). The gels for band shift assays were prepared using either 1x TA or 1x TB buffer, respectively. Depending on the size of the samples that were to be separated, different formulations of gels were used, varying in the weight

percentage of added agarose. For dsDNA and ssRNA samples with sizes above 1000 bp or 1000 nt, respectively, gels with one percent of agarose were used. All samples below this size, down to 250 nt for ssRNA, were separated on 1.5 % agarose gels. ssRNAs with a size below 250 nt were separated on 2 % agarose gels. In addition, agarose gels for the separation of DNA were pre-stained with SYBR Safe in 1x concentration. Additionally, short ssRNAs were also run on 10 % TBU polyacrylamide gels. These gels were prepared from 1x TBE buffer, mixed with urea to a final concentration of 7 M and 10 % Rotiphorese (Roth) with acrylamide/bisacrylamide in a ratio of 29:1. The polymerization of the gel stock solution was started by addition of 0.01 volume of 10 % APS solution and 0.001 volumes of TEMED.
In general, for the preparation of DNA samples, the DNA was mixed with 6x purple gel loading dye (NEB, 2.5 % Ficoll-400, 10 mM EDTA, 3.3 mM Tris·HCl, 0.08 % SDS, 0.02 % Dye 1, 0.001 % Dye 2, pH 8 at 25°C). In case of RNA, samples were usually mixed with 2x RNA loading dye (Thermo Scientific, 95 % Formamide, 0.025 % SDS, 0.025 % Bromophenol Blue, 0.025 % Xylene Cyanol FF, 0.025 % Ethidium Bromide, 0.5 mM EDTA), and heated up to 70°C for five minutes. All samples for band shift assays were prepared using the corresponding EMSA buffer (see Table 4).
Agarose gels were run at room temperature with 6.5 V/cm for 45 minutes, in the case of general gel electrophoresis, or, respectively, cooled with 2 V/cm for three hours in the case of band shift assays. Polyacrylamide gels were pre-run for 30 minutes at 10 W, and for another 90 minutes at 10 W after sample loading and stained with 1x SYBR Gold dye (Invitrogen) in 1x TBE for five minutes prior to scanning. For size comparison, samples were run alongside GeneRuler 1 kb Plus DNA Ladder (Thermo Scientific) in case of DNA samples, RiboRuler High Range RNA Ladder (Thermo Scientific) for RNAs longer, and RiboRuler Low Range RNA Ladder (Thermo Scientific) for RNAs shorter than 1000 nt.

2.3.2 Gel electrophoresis of proteins

For analysis of protein preparations, samples were separated on SDS containing polyacrylamide gels, using a general formulation for Lämmli gels (see Table 4). The stock solutions for stacking (StaB) and resolving (ResB-X) gel were prepared using Rotiphorese (Roth) with acrylamide/bisacrylamide in a ratio of 37.5:1. The final concentration of acrylamide in the resolving gels was 10 % for proteins above 100 kDA in size, and 12.5 % for all other samples, respectively. The preparation of gels was done by starting polymerization of acrylamide by addition of 0.01 volume of 10 % APS solution and 0.001 volumes of TEMED, with the stacking gel being poured on top of already polymerized resolving gel. Prior to loading, samples were mixed with 2x SDS Loading Buffer (100 mM Tris·HCl pH 6.8, 4 % SDS, 0.2 % Bromophenol Blue, 20 % Glycerol, 200 mM DTT) and heated up to 95°C for five minutes. The gels were run in 1x SDS-RB with 15 V/cm for 90 minutes at room temperature. For size

estimation, the samples were run alongside BlueClassic Prestained Protein Marker (Jena Bioscience) as a size standard. To visualize protein bands, the gels were stained with InstantBlue Protein Stain (Expedeon) for at least 30 minutes to overnight and washed with H_2O before imaging, to improve the contrast.

2.3.3 Chemically competent cell preparation and transformation

Chemically competent cells for transformation by heat shock were prepared by growing cells in LB media to an OD_{600} between 0.4 and 0.6. The cells were cooled down on ice, harvested by centrifugation for ten minutes at 4°C and 4000 rcf, the pellet was resuspended in 40 mL ice cold TB1 per one liter culture, spun down again and finally resuspended in 2 mL ice cold TB2. The suspension was split into 50 µL aliquots, which were subsequently flash frozen in liquid nitrogen and stored at -80°C until usage.
For the actual transformation, cells were allowed to thaw on ice, mixed with 1 ng of the desired DNA, incubated on ice for 30 minutes. After a heat-shock for 30 seconds at 42°C, cells were incubated for another five minutes on ice, before being mixed with 950 µL pre-warmed LB. Following a recovery incubation at 30°C to 37°C with rapid shaking (900 rpm), to allow for proper expression of resistance markers, cells were plated on 1.5 % LB-plates and incubated at the same temperature overnight.

2.3.4 Electro-competent cell preparation and transformation

The protocol for preparing competent cells by electroporation is based on a protocol established by *Tu et al*. [189]. In short, cells were grown LB media to an OD_{600} between 0.1 and 0.2. Per transformation, 1.4 mL culture were centrifuged at 25°C and 9000 rcf for one minute. The cell pellet was resuspended in 1 mL H_2O and spun down again. In total, cells were washed three times, before finally being resuspended 50 µL H_2O. The suspension was mixed with the desired nucleic acid and electroporated in an Eporator (Eppendorf) with 12500 V/cm. Immediately after electroporation, cells were mixed with 950 µL pre-warmed LB media and incubated at 30°C to 37°C under vigorous shaking (900 rpm). Subsequently, cells were either directly plated or used for downstream experiments.

2.3.5 PCR and DNA clean-up

In general, PCRs were performed using Q5 Hot Start High-Fidelity 2x Master Mix (NEB), following the manufacturer's recommendations for the PCR cycling program and annealing temperatures for primers in a final concentration of 500 nM were calculated using the Tm Calculator version 1.13.0 (NEB). The respective PCR template plasmids were usually provided as 1 ng per 50 µL reaction mix. Following the PCR, the template plasmid was digested by adding 1 µL DpnI (NEB) and incubation the mix at 37°C for one hour. Using the Monarch PCR

& DNA Clean-Up Kit (NEB), the amplicon was subsequently purified, using nuclease-free H_2O for the elution.

Exceptions to this protocol were made for all amplicons based on the (-)-strand of MS2(wt), specifically $[F30\text{-}Bro(+/-)]_{UTR(-)}$ and $[F30\text{-}Bro(+/-)]_{MS2(-)}$, as well as all transposition cassettes based on the MuA transposon. These amplicons were synthesized using GoTaq G2 Hot Start Master Mix (Promega), following the supplier's instructions with regards to PCR program and elongation times. In these cases, the annealing temperatures were calculated using Tm for Oligos Calculator (Promega). All other parameters and subsequent steps were identical to the ones described above.

The most important primer and amplicon sequences, that were not exclusively used as IVT templates, are listed in tables Table 15 and Table 16, respectively.

2.3.6 IVT and RNA clean-up

In vitro transcription reactions were performed by using MEGAscript T7 Transcription Kit (Invitrogen), or, respectively, MEGAshortscript T7 Transcription Kit (Invitrogen), if the desired RNAs were shorter than 500 nt. Thereby, reaction mixes were prepared as recommended by the supplier and templates were added with 500 ng per 20 µL reaction. The reactions were incubated at 37°C for two hours, and after addition of 1 µL TURBO DNAse (Invitrogen), incubated for another 30 minutes at 37°C.

Finally, the newly transcribed RNAs were purified using the Monarch RNA Clean Up Kit (NEB), following the manufacturer's instructions, and eluting the RNAs in nuclease-free H_2O.

The sequences of all relevant IVT templates are listed in Table 17

2.3.7 Plasmid preparation

For the isolation of plasmids from bacterial cells, a colony of freshly transformed cells was picked and incubated overnight at 30°C in 5 mL LB, supplemented with the corresponding selection marker. After harvesting the cells by centrifugation for ten minutes at room temperature and 3200 rcf, the plasmid DNA was isolated using the NucleoSpin Plasmid Mini Kit (Macherey & Nagel), following the suggested protocols for either high or low copy number plasmids and eluting the DNA in nuclease-free H_2O. The correct sequences of the isolated plasmids were confirmed by Sanger sequencing by Microsynth Seqlab GmbH (Göttingen, Germany), whereby sequencing primers were designed to be approximately 1000 bp apart.

2.3.8 Determination of concentrations

The concentrations of DNA, RNA and protein preparations were determined using the absorbances at 260 nm for RNA and DNA, and at 280 nm for proteins, respectively. The absorbances were measured on a NanoDrop One (Thermo Scientific) and in the case of DNA

concentrations were directly taken from the device (1 A_{260} corresponds to 40 ng/µL). In the case of RNA and proteins, coefficients were calculated by using Oligo Calc [190] for RNA and ProtParam [191] for proteins, respectively. Subsequently, the concentrations were calculated using the correlation between the absorbance A_λ at wavelength λ, the absorption coefficient ε_λ ($Mol^{-1}\ cm^{-1}$), the optical pathlength d (cm) and the concentration c ($mol\ L^{-1}$) described by the Lambert-Beer law shown below.

$$\boldsymbol{c = \frac{A_\lambda}{\varepsilon_\lambda \cdot d}}$$

2.4 Mutagenic IVT

The libraries of potential MP binding RNAs were generated through mutagenic IVT, as described by *Pezo et al.* [192]. To introduce mutations into the hairpin RNAs during the IVT, the T7 reaction mixes were supplemented with 5 mM $MnCl_2$ and biased NTP pools. These biased pools (AU_{GC}, AC_{GU}, GU_{AC}, GC_{AU}) were prepared by mixing the corresponding NTPs in a ratio of 100:1 (X_Y = 100 X to 1 Y), leading to a final concentration of 5 mM and 50 µM, respectively. Following incubation at 37°C for four hours, the corresponding RNAs from each biased NTP pool were combined, the DNA template digested, and the final libraries purified. The libraries, as well as reference RNAs from standard IVT, were sequenced by Illumina sequencing in the core facility at the Max Planck Institute of Biochemistry (Martinsried, Germany).

2.5 Genomic knock-ins

The protocol to insert the gene *sacB* into the *rnc* gene within genomes of *E. coli* HMS-174(DE3) is based on the Quick & Easy *E. coli* Gene Deletion Kit (Gene Bridges), which utilizes the Red/ET system for homologous recombination of a cassette into a target locus. In general, the procedure is identical to the manufacturer's instructions, however, some adaptations were made. The sequence for the flanking regions, as well as the *rnc* coding sequence itself were taken from a reference genome of *E. coli* K12 (NCBI accession ID U00096.3).

One modification was to extend the original cassette used in the kit to also include a gene of interest, as well as a constitutive promoter for expression of this gene. Therefore, two plasmids were cloned. The first, pGEM-T_FRT-KanR-P70a, encodes for the 5'-end of the original cassette, but now also includes the sequence for the P70a promoter downstream of the Kanamycin resistance marker [193,194]. The other one, pBAD33_sacB-3'-FRT, contains the coding sequences for *sacB* (NCBI Gene ID 936413), integrated into a pBAD33 backbone and followed by the 3'-end of the original cassette. To assemble the full cassette, the corresponding fragments were first amplified by PCR from their respective plasmids, and finally assembled into the final cassette by an extension PCR with Q5 polymerase and both fragments in a final concentration of 10 ng/µL, following the program shown below:

95°C	60 seconds	
95°C	10 seconds	¯|
65°C	30 seconds	| 10 cycles
68°C	10 minutes	_|
68°C	10 minutes	
15°C	hold	

Following the PCR, the desired product was isolated through gel extraction. The second modification was the preparation of electrocompetent cells for transformation of both the Red/ET encoding plasmid, as well as the final cassettes. Here, electro competence was established by following the protocol described in section 2.3.4.
Lastly, successful genomic integration was confirmed through selection for Kanamycin resistance (25 µg/mL Kanamycin), colony PCR and Sanger sequencing of amplicons of the modified genomic *rnc* locus. The sequences for the cassette fragments and the PCR primers are shown in tables Table 15 and Table 16, respectively.

2.6 Cloning and reprogramming of EcI5 introns

2.6.1 EcI5 components and reprogrammed plasmids

The design of the vectors encoding for the EcI5 intron without open reading frame (EcI5(ΔORF)), followed by the CDS for the EcI5 maturase, is based on findings by *Zhuang et al.* [195]. The full construct was assembled from three gene blocks (IDT) and a pBAD33 backbone, using NEBuilder HiFi DNA Assembly master mix (NEB), with the intron being programmed to insert into the ORF of *lacZ* of *E. coli* K12 strain at position 1806 in the sense strand (1806s) [195]. This yielded the vector pBAD33_EcI5_lacZ-1806s. The full sequence of the EcI5(ΔORF) intron is shown in Table 16, albeit with generic sequences for insertion programming (depicted as Ns). The CDS of the maturase was taken from the NCBI repository (NC_002128.1 (36274 – 37998)) [196]. The insertion into the *lacZ* gene at position 1806s requires the nucleotides for exon binding site 3 (EBS3) and intron binding site 3 (IBS3) to be A and T, respectively. In a first step, to also enable retargeting to other loci, pBAD33_EcI5_lacZ-1806s was modified in both positions (EBS3 and IBS3) to generate the other three combinations as well. This was done through two separate PCRs with mutagenic primers, each followed by ligation and transformation. As a result, the plasmids pBAD33_EcI5-G (EBS3 = G, IBS3 = C), pBAD33_EcI5-C (EBS3 = C, IBS3 = G), pBAD33_EcI5-T (EBS3 = T, IBS3 = A) were obtained. Finally, to reprogram the insertion site, two more PCRs with loci specific primers, each coupled to ligation and transformation, were performed. Primer sequences are shown in Table 15, with the locus specific parts depicted as generic nucleotides (N). The method to determine valid insertion sites is described in the following section.

2.6.2 Reprogramming algorithm

The algorithm to find suitable insertion sites for the EcI5 intron is based on the algorithm written for the related group II intron LtrB but using sequence requirements and nucleotide frequencies reported for the EcI5 intron [195,197].

In general, the algorithm calculates a score for a specific site based on the sum of individual scores over a stretch of 36 nt, consisting of 26 nt before (-26 to -1) and ten nt after the actual insertion site (+1 to 10). For each position, the individual score for a specific nucleotide is calculated as the binary logarithm of the frequency ratio between the selected library and the initial pool library r(i). With n being the maximum number of positions (36) within the analysed stretch and i being the current nucleotide's position, the total score S is calculated as:

$$S = \sum_{i}^{n} log_2(r(i))$$

The ratio for a given position i (r(i)) itself is calculated as shown below, with the frequencies $f(i)_s$ and $f(i)_p$ for the selected library, or respectively, the pool library, being calculated as the ratio between the count of a specific nucleotide $n(i)_x$ over the total count of all nucleotides $t(i)_x$ sequenced at the given position.

$$r(i) = \frac{f(i)_s}{f(i)_p} = \frac{\frac{n(i)_s}{t(i)_s}}{\frac{n(i)_p}{t(i)_p}}$$

The frequency ratios derived from *Zhuang et al.* [195] are shown in Table 5, rounded to three decimals. For position -7, the ratios are the arithmetic mean of the selections for IBS1/EBS1 and IBS2/EBS2. As reported by *Zhuang et al.* [195], the positions -26 to -14 and +3 to +10 in the target locus are required for binding of the maturase, positions -13 to -8, -6 to -1 and +1 for binding by EBS2, EBS1 and EBS3, respectively [195]. The interaction between target IBS2 and intron EBS2 favours no base pairing at position -8, the interaction between IBS1 and EBS1 however can tolerate either case at position -6 and EBS3 again requires base pairing with IBS1 at position +1.

A modified version of the frequency ratios depicted in Table 5 can also be used for calculation of the score. The modifications are thought to reduce a potential training bias due to limited pool size and allow for a less restricted sequence space for the DNA-RNA base pairing. This is achieved by setting the ratios for positions -13 to -9 and -5 to +1 to a value of 1.000, which results in no contribution of these positions to the calculation of the final score [195]. Thus, the theoretical maximum range for calculated scores using the tight restrictions is -48.45 to 29.46, and for the looser restrictions it is -29.55 to 21.39.

Table 5: Nucleotide frequency ratios for EcI5 reprogramming.

Position	*A*	*C*	*G*	*T*
- 26	0,760	1,500	0,900	1,200
- 25	0,864	1,000	1,051	1,000
- 24	1,036	1,000	0,902	1,167
- 23	1,091	0,692	1,041	0,941
- 22	1,474	0,450	1,106	0,733
- 21	1,037	1,125	1,077	0,684
- 20	1,100	1,353	0,760	1,214
- 19	1,353	0,500	1,244	0,579
- 18	0,250	4,875	1,133	0,150
- 17	0,458	4,923	0,440	0,214
- 16	0,667	1,143	1,023	1,231
- 15	4,050	0,050	0,326	0,200
- 14	3,846	0,000	0,000	0,000
- 13	**0,192**	**1,857**	**1,412**	**0,808**
- 12	**0,750**	**0,278**	**1,262**	**1,250**
- 11	**0,294**	**0,158**	**1,818**	**0,600**
- 10	**0,067**	**0,333**	**1,933**	**0,263**
- 9	**0,125**	**0,100**	**1,457**	**1,611**
- 8	0,632	0,056	1,854	0,524
- 7	0,689	1,396	1,628	1,714
- 6	0,900	0,605	3,308	0,552
- 5	**0,452**	**1,710**	**0,650**	**1,105**
- 4	**0,968**	**1,111**	**0,947**	**0,957**
- 3	**1,269**	**0,625**	**1,615**	**1,000**
- 2	**1,941**	**0,707**	**0,833**	**0,958**
- 1	**0,227**	**1,571**	**0,353**	**1,308**
+ 1	**2,158**	**1,682**	**0,481**	**0,313**
+ 2	1,900	0,525	0,682	1,368
+ 3	0,769	1,465	0,200	0,800
+ 4	0,870	1,091	0,286	1,400
+ 5	0,625	0,150	0,250	2,759
+ 6	1,462	0,646	2,545	0,759
+ 7	1,000	1,000	1,176	0,864
+ 8	1,000	1,158	0,864	0,800
+ 9	0,826	1,103	1,214	0,840
+10	0,700	0,905	0,714	2,067

In general, the code will scan any input sequence (sense) in increments of 36 nt, starting at the 5'-end and sliding the increment in steps of one nucleotide towards the 3'-end, while also scanning the complementary sequence (anti) in the same way. Additionally, in case of coding sequences, the full algorithm will also find and translate the CDS into a degenerate amino acid (AA) sequence. Subsequently, this sequence will get fragmented into 13 AA long increments with a slide between each increment of one AA. Each of these increments will be reverse translated into all possible combinations of synonymous codons, resulting in new pool of 39 nt long sequences. These will then again be scanned for all three reading frames. This procedure is then repeated for the complementary strands of each of these 39 nt long stretches.

Each hit with a score above a set threshold will be stored in a file, including score, sequence,

position of insertion, IBS1 – IBS3, EBS1 – EBS3, as well as the required plasmid and primers to clone the desired EcI5 variant. Finally, the final output file will only contain the hits with the highest score for each distinct position, while the additional hits are stored in a separate file. Furthermore, the allowed codons for the codon permutation can be restricted to a set number of the highest frequency codons, based on codon frequencies of *E. coli* strain K12 derived from the Codon Usage Database [198].
The full, executable code is shown in section 7.5, including a documentation of the individual tasks performed by each block of code.

2.7 *In vivo* application of EcI5

2.7.1 Knock-out of lacZ

To verify *in vivo* disruption of the *lacZ* gene of *E. coli* HMS-174(DE3), cells were transformed with the plasmids pBAD33_EcI5(ΔORF)_lacZ-1806s and pBAD33_EcI5(ΔORF)_sacB-491s, respectively. Starter cultures of transformed cells were then grown in LB(Cam) to an OD_{600} of 0.1, induced with 0.5 mM IPTG, grown at 37°C for three hours, plated on 1.5% LB(Cam)-agar plates with 0.1 mg/mL X-Gal and 0.1 mM IPTG and finally incubated overnight at 37°C. For further controls, additional cultures were treated equally but not induced.

2.7.2 Knock-out of sacB

For the identification of suitable target sites in the *sacB* gene, the coding sequence was analysed using the algorithm described in section 2.6.2, with the following parameters: loose restrictions, no amino acid shuffling, threshold score >7.5. Thus, the three target sites at positions 116, 309 and 491 in the sense strand were identified and consequentially the corresponding plasmids were cloned. The EBS1 parts were designed to form 5 base pairs with the respective IBS1. Transformation of these plasmids into *E. coli* HMS-174(DE3)^{sacB} was followed by inoculation of starter cultures. The cultures were grown at 30°C to an OD_{600} of 0.1, induced with 0.1 mM IPTG, grown at 37°C for three hours and finally approximately $5 \cdot 10^3$ (100 µL of $OD_{600} = 10^{-4}$) cells were plated on 1.5% LB(Cam)-agar plates supplemented with 5% sucrose. The respective controls were identically treated, however not induced with IPTG. The plates were grown at 37°C overnight. Cells transformed with the pBAD33_EcI5(ΔORF)_lacZ-1806s plasmid were used as negative control.

2.8 Cloning and purification of MP

2.8.1 Reverse transcription and cloning of wildtype mp gene

An amplicon of the MS2 maturation protein (MP) was generated by reverse transcription from MS2 RNA (Roche), followed by several PCR cycles, using primers that corresponded to the 5'-end of the coding sequence and the 3'-end complementary sequence, respectively, but

changing the first codon from GTG (V) to ATG (M). Both steps were performed using the OneTaq One-Step RT-PCR Kit (NEB) with the following program:

48°C	30 seconds	
94°C	1 minute	
94°C	15 seconds	¯|
54°C	30 seconds	| 40 cycles
68°C	2 minutes	_|
68°C	5 minutes	
15°C	hold	

After the RT-PCR, the amplicon was purified and ligated into an empty cloning vector, using the pGEM-T Easy Vector system (Promega) and incubating the ligation reaction at room temperature for one hour. From this mix, 2 µL were transformed into chemically competent *E. coli* Top10 cells. Subsequently, these cells were plated on LB(Carb)-plates containing IPTG (1 mM) and X-Gal (0.2 mg/mL), allowing for blue/white screening to identify the desired transformants. The MP sequences in isolated plasmids were confirmed by Sanger sequencing and subsequent comparison to MS2 reference genomes (NCBI GenBank ID: GQ153925.1, MT830620.1, EF204940.1, GQ153924.1, GQ153927.1) [199,200], leading to the identification of several clones harbouring intact MP coding sequences either in sense (pGEM-T Easy_MP) or antisense orientation (pGEM-T Easy_MP_inverse).

2.8.2 Cloning of MP expression plasmids

In general, all plasmids for the expression and subsequent isolation of tagged MP were prepared by assembling the corresponding fragments, amplified by PCR, using NEBuilder HiFi DNA Assembly Master Mix (NEB), which was followed by transformation of the assembly mix into chemically competent *E. coli* Top10 cells.

For the expression of hexa-histidine tagged (His_6) MP in *E. coli* BL21(DE3) cells, the coding sequence of the MP was cloned into a pET28a backbone, downstream of a T7 promoter, resulting in the two plasmids pET28a_His_6-MP and pET28a_MP-His_6. The plasmid for the amplification of the pET28a backbone was kindly provided by the Schwille group at the Max Planck Institute of Biochemistry (Martinsried, Germany).

All constructs of the MP with the solubility enhancing maltose binding protein tag (MBP), were assembled using a plasmid encoding the MBP-tag for amplification of the required fragment for HiFi assembly. This plasmid was kindly provided by the core facility at the Max Planck Institute of Biochemistry (Martinsried, Germany). These constructs include pET28a_MP-MBP-His_7 and pET28a_His_6-MBP-MP.

The mutant variations, as well as variants of the recombinant MP tagged with a twin

streptavidin tag (TST) [201] were cloned by PCR with modified primers, harbouring the modified sequences as 5'-overhangs, followed by ligation of linear amplicons using KLD Enzyme Mix (NEB) and transformation into chemically competent *E. coli* Top10 cells. The plasmids pET28a_MP-TST-MBP-His_6, as well as pET28a_MP(mut)-His_6, pET28a_MP(1a)-His_6, pET28a_MP(1b)-His_6, pET28a_MP(1a2)-His_6, pET28a_MP(1b2)-His_6, pET28a_MP(1a3)-His_6 and pET28a_MP(1b3)-His_6 were obtained from this procedure.
The plasmid pBAD33_CBD-MP-MBP-His_7 encodes for a chitin binding domain (CBD) embedded in the self-cleaving *Sce* VMA intein [202], both taken from the pTYB21 vector (NEB), followed by the MP-fusion construct from the pET28a_MP-MBP-His_7 vector. Both parts were assembled into a pBAD33 backbone, allowing for induction of protein overexpression with L-Arabinose. The plasmid for the PCR of the pBAD33 backbone was kindly provided by L.I. Weise.

2.8.3 Expression and purification of MP

The expression of various MP-constructs under *in vitro* conditions was achieved by programming the cell free translational system PURExpress (NEB) with plasmids encoding for the corresponding construct. In detail, 5 µL Solution A were mixed with 3.75 µL Solution B and brought to a final volume of 12.5 µL with H_2O and the desired plasmid in a final amount of 150 ng. Subsequently, the reaction mixes were incubated at 37°C for three hours.

N-terminally tagged His_6-MP and MBP-MP were isolated as followed. BL21(DE3) were transformed with the corresponding plasmids. A single colony per construct was used for inoculation of a starter culture in LB(Kan), grown overnight at 30°C. With this starter, 25 mL LB(Kan) were inoculated and grown to an OD_{600} around 0.5 at 37°C. Following induction with 1 mM IPTG, cells were grown at 16°C overnight. Cells were pelleted and resuspended in 1 mL lysis buffer (50 mM Tris pH 7.5, 250 mM NaCl, 1 mM DTT) and sonicated on ice. After pelleting the cell debris at 4°C with 25,000 g for 30 minutes, the supernatant was discarded, and the pellet washed three times with 1 mL lysis buffer. The pellet was denatured by addition of 500 µL of the tested denaturing buffers (Table 11) and incubation at 37°C for one hour, shaking at 1000 rpm. Non-dissolved proteins were pelleted by centrifugation at room temperature with 20,000g for 45 minutes. The supernatants from pellets denatured with denaturing buffer 2 were subjected to refolding by dialysis against refolding buffer (50 mM Tris pH 7.5, 100 mM NaCl, 1% sodium polyacrylate, 0.5 M Urea) overnight at 4°C. Subsequently, insoluble aggregates were pelleted by centrifugation at 4°C with 20000 g for one hour. Finally, the supernatant was applied to 50 µL equilibrated (refolding buffer) HisPur™ Ni-NTA resin (Thermo Fisher Scientific). The resin was washed three times with five column volumes (CV) refolding buffer and the bound MP constructs eluted with three CV elution buffer (50 mM Tris pH 7.5, 100 mM

NaCl, 1% sodium polyacrylate, 0.5 M Urea, 300 mM Imidazol, 10% glycerol). The final preparation was flash frozen in liquid nitrogen and stored at -80°C until usage.

For large scale expression and purification of MP-constructs with a CBD-tag, *E coli* Top10 cells were transformed with the pBAD33_CBD-MP-MBP-His_7 vector. From a single colony, a 5 mL starter culture in LB(Cam) was prepared and incubated at 28°C overnight. This starter was used to inoculate one litre of fresh LB(Cam), the culture was grown at 28°C to an OD_{600} of 0.5, overexpression induced by addition of L-Arabinose in a final concentration of 0.2 %, and the cells subsequently incubated at 16°C overnight. Following harvesting of the cells at 4°C and 6,000 rcf for 45 minutes, the cell pellet was resuspended in 40 mL LyB1 and lysed by sonication in an ice/water bath. After pelleting cell debris by centrifugation at 4°C and 10,000 rcf for 45 minutes, the clear supernatant was incubated with 2.5 mL of pre-equilibrated (WBL1), packed Chitin resin (NEB) for 45 minutes at 4°C. The resin was then packed into a gravity flow column, washed three times with 20 CV WBL1, and flushed with five CV EB1. Another three CV EB1 were added, and the column incubated 72 hours at 4°C under gentle shaking, to allow the cleavage of the intein to take place. The cleaved MP was washed of the column with additional three times three CV EB1 and up concentrated in 20 minutes increments at 4°C and 3,200 rcf, using a Amicon Ultra-15 spin concentrator (Merck) with a molecular weight cut off at 30 kDa. Finally, the buffer was exchanged with DB1, and the final preparation flash frozen in liquid nitrogen and stored at -80°C until usage.

2.9 Plaque assays

In general, plaque assays were used to detect infection of *E. coli* strain F^+5695 with MS2 bacteriophage. This was achieved by firstly inoculating fresh LB(Strep) with a fraction of an overnight culture, growing this culture at 30°C to an OD_{600} of 1.0 and finally incubating 100 µL of this culture with the infectious agent at room temperature for 30 minutes. In the case of electroporation of MS2 RNA, the protocol until infection was modified by preparing electrocompetent F^+5695 cells (1.4 mL, OD_{600} 0.13), electro-transforming MS2 RNA into the cells, adding 900 µL pre-warmed LB medium and incubating them for 30 at 37°C and shaking with 450 rpm.

Subsequently, the cell suspension was mixed with 3 mL 0.5 % LB-agar (Strep, 45°C), or respectively, in case of electroporated MS2 RNA 2 mL 0.75 % LB-agar (Strep, 45°C) and poured on pre-warmed 1.5 % LB(Strep)-agar plates. After letting the gel solidify at room temperature for 15 minutes, the plates were incubated at 37°C overnight.

Three different methods were used to prepare the infectious agent. The first method was to program a 12.5 µL PURExpress reaction with MS2 RNA, adding PEG 6000 to 0.3 %, and incubate this reaction mix at 37°C for 2.5 hours. The second method was to prepare an identical reaction, but to program it with a plasmid encoding for the MP. Addition of the MS2

RNA followed an initial incubation of this reaction mix at 37°C for 3 hours, followed by a final incubation at room temperature for 15 minutes to allow formation of minimal infectious units. The last method was to incubate MS2 RNA with purified MP-MBP-His_7 in 20 µL of 1x PBB for 15 minutes at room temperature, followed by incubation on ice until the infection.

2.10 Cloning of MS2 variants

All variants of the bacteriophage MS2 that were used for this thesis are derivatives of the pGEM-T Easy_MS2(wt) vector, encoding for an infectious MS2 wildtype, cloned by S.V. Mayr. Both plasmids encoding for MS2 constructs, with an inserted F30 stem and Broccoli aptamer (pGEM-T Easy_MS2(Bro+) and pGEM-T Easy_MS2(Bro-)), were provided by L.I. Weise [180,182]. The variant MS2(ddef), a version with disrupted start codon of the lysis gene (M1T), as well as a loss-of-function mutation of the coat protein (V75E, A81G) [203], was produced from the pGEM-T Easy_MS2(wt) vector in two sequential cloning steps through PCRs with modified primers. The resulting pGEM-T Easy_MS2(ddef) vector was then used, in an identical procedure, to generate the pGEM-T_ApaLI_MS2(ddef) vector. This version of the predecessor plasmid contains ApaLI cleavages sites on each end of the MS2 genomic sequence, as well as a shortened backbone, to reduce background integration during the generation of the random insertion libraries. In addition, the MS2(qdef) variant, which harbours the same mutations as MS2(ddef), as well as a disrupted start codon for the MP (M1A) and a defective MS2rep subunit (D341S, D342V) [180,204] was generated by PCR with mutagenic primers, starting from pGEM-T_ApaLI_MS2(ddef) vector.

2.11 Cloning and purification of MuA and insertion cassette

2.11.1 Cloning of TnP-CamR & Tnp-KanR

For the random insertion by MuA transposase of the resistance genes against Kanamycin and Chloramphenicol, respectively, into MS2 genomes, two plasmids were generated that encoded each for one of the two selection markers embedded in a minimal Mu phage transposon scaffold. In a first step, a gene block (Integrated DNA Technologies, IDT) encoding for the empty cassette derived from the commercially available Mutation Generation Kit (Thermo Scientific), but with wildtype MuA recognition sites R1/R2 and cleavage sites for BglII [205,206], was inserted into a shortened pGEM-T backbone (pMini_Tnp) using NEBuilder HiFi DNA Assembly master mix (NEB). Subsequently, also using HiFi assembly, the coding sequences for the resistance genes were inserted into the empty transposon, resulting in the two plasmids pMini_Tnp-KanR and pMini_Tnp-CamR, respectively. The sequence of the empty transposon is shown in Table 16.

2.11.2 Cloning of MuA expression plasmid

The coding sequence of hyperactive MuA (W160R, E233K, W345R) transposase is based on findings of *T. Rasila et al.* [207]. A gene block (IDT) encoding the CDS was cloned after the self-cleaving *Sce* VMA intein with CBD, into the pBAD33 backbone also used for overexpression of the MS2 MP. The resulting expression vector was pBAD33_CBD-haMuA.

2.11.3 Purification of hyperactive MuA

For the expression and purification of hyperactive MuA transposase, *E coli* Top10 cells were transformed with the pBAD33_CBD-haMuA vector. From a single colony, a 5 mL starter culture in LB(Cam) was prepared and incubated at 28°C overnight. This starter was used to inoculate one litre of fresh LB(Cam), the culture was grown at 28°C to an OD_{600} of 0.5, overexpression induced by addition of L-Arabinose in a final concentration of 0.2 %, and the cells subsequently incubated at 16°C overnight. Following harvesting of the cells at 4°C and 6,000 rcf for one hour, the cell pellets were resuspended in 40 mL LyB3 and lysed by sonication in an ice/water bath. After pelleting the cell debris by centrifugation at 4°C and 10,000 rcf for 30 minutes, the clear supernatant was incubated with 2.5 mL of pre-equilibrated (WBL3), packed Chitin resin (NEB) for 30 minutes at 4°C. The resin was then packed into a gravity flow column, washed two times with five CV WBL3, then four times five CV WBH3, and again with three times five CV WBL3. After being flushed with five CV EB3, another three CV EB3 were added and the column incubated 72 hours at 8°C and under gentle shaking, to allow the cleavage of the intein to take place. The cleaved transposase was washed of the column with additional three times three CV EB3 and up concentrated in several 20 minutes increments at 4°C and 3,200 rcf, using a Amicon Ultra-15 spin concentrator (Merck) with a molecular weight cut off at 30 kDa. After buffer exchange to DB3, the protein was flash frozen in liquid nitrogen and stored at -80°C until usage.

2.12 MuA library generation

The first step to generate libraries of MS2 variants with randomly inserted selection marker cassettes, was the preparation of the corresponding transposon cassettes. This was done by amplification of these components by PCR, followed by digest of the PCR template with DpnI for one hour at 37°C and clean-up of the amplicons. Afterwards, the ends of the insertion cassette amplicons were processed by digestion with BglII (NEB) for 30 minutes at 37°C, followed by another clean-up step.

The next step was the insertion of the cassettes into the target plasmids encoding for the corresponding MS2 variants. This was achieved by mixing 400 ng target and 200 ng cassette in 1x RIMB, addition of MuA transposase in a final concentration of 0.22 µM, followed by incubation at 30°C for two hours. After a clean-up step, the modified plasmids were screened

for successful integration by transforming the DNA into electrocompetent *E. coli* Top10 cells, incubation at 37°C for one hour in 1 mL LB, and subsequent selection in 20 mL liquid culture with antibiotics for the resistance markers in the backbone (Carbenicillin, 50 µg/mL) and the cassette (Kanamycin, 25 µg/mL / Chloramphenicol, 17 µg/mL). Following an overnight incubation of these cultures at 30°C, the primary libraries were prepared by plasmid isolation. Finally, a PCR was performed on the primary library plasmids to generate templates for IVT, followed by IVT of the insertion mutant RNAs with T7 polymerase. For transformation of these RNAs, *E. coli* F^+5695 were made electrocompetent and electroporated with 2.5 µg of the RNAs derived from MS2(ddef) and MS2(qdef) plasmids. Cells were subsequently plated on plates with the respective antibiotic in concentration ranging from 5 µg/mL to 50 µg/mL for Kanamycin and 5 µg/mL to 34 µg/mL for Chloramphenicol, respectively, and grown overnight at 37°C.

2.13 Cloning and purification of MS2rep, Qßrep and co-factors

2.13.1 Cloning of MS2rep, Qß, EF-Tu and EF-Ts expression plasmids

The expression vectors encoding for the replicase subunits of the bacteriophages MS2 (MS2rep) and Qβ (Qβrep), both with N-terminal and C-terminal His_6-tag were cloned from the corresponding predecessor plasmids pBAD33_TuTs-MS2rep and pBAD33_TuTs-Qβrep, respectively, which were provided by L.I. Weise. After PCR with mutagenic primers, followed by ligation of the linear product with KLD Enzyme Mix (NEB) and subsequent transformation, the plasmids pBAD33_His_6-MS2rep, pBAD33_ MS2rep-His_6, pBAD33_His_6-Qβrep and pBAD33_Qβrep-His_6 were obtained. The vectors pBAD33_EF-Tu-His_6 and pBAD33_EF-Ts-His_6 were prepared using the same predecessor plasmids and adapted mmutagenic primers.

2.13.2 Cloning of additional co-factor expression plasmids

The coding sequences for the expression plasmids pBAD33_S1-His6, pBAD33_His_6-IF1 and pBAD33_His_6-IF3 were obtained by colony PCR from *E. coli* Top10 cells of the respective coding sequences of ribosomal protein S1 and the translation initiation factors IF1 and IF3, using modified gene specific primers to introduce the His_6-tag. Following this PCR step, the amplicons were assembled into a pBAD33 vector backbone through standard HiFi assembly, using NEBuilder HiFi DNA Assembly Master Mix (NEB).

The vectors for overexpression of methionyl-tRNA formyltransferase (MTF), alanine-tRNA ligase (AlaRS), asparagine-tRNA ligase (AsnRS), isoleucin-tRNA ligase (IleRS) and both phenylalanine-tRNA ligase subunits (PheRSα, PheRSβ) were assembled in an identical fashion. However, in these cases the coding sequences were amplified from the pLD2 expression plasmid, provided by K. Libicher. This yielded the expression vectors pBAD33_MTF-His_6, _His_6-AlaRS, _His_6-AsnRS, _IleRS-His_6 and _His_6-PheRS(α+β).

2.13.3 Expression and purification of replicases and co-factors

The pBAD33 based expression vectors were transformed into chemically competent *E. coli* Top10 cells, a single colony per construct was grown overnight in 5 mL LB(Cam) at 30°C and subsequently used to inoculate one litre of fresh LB(Cam). This culture was then grown to an OD_{600} of 0.5 and overexpression was induced by addition of L-Arabinose in a final concentration of 0.2%. The induced cultures were grown at 16°C overnight, harvested by centrifugation at 4°C and 5,000 rcf for one hour, before being resuspended in 40 mL LyB2. The lysis by sonication in an ice/water bath was followed by another centrifugation step at 4°C and 10,000 rcf for 30 minutes, to pellet the cell debris. The clear supernatants were applied to 5 mL packed and equilibrated (WBL2) HisPur NiNTA resin (Thermo Scientific) and incubated at 4°C for 30 minutes. Subsequently, the resins were loaded into empty gravity columns. In case of the individual co-factors, except S1, EF-Tu, EF-Ts, IF1 and IF3, the resin was washed five times with five CV WBL2. In case of the replicase subunits and the co-factors listed above, the resin was washed twice with five CV WBL2, then twice with five CV WBH2, and again twice with five CV WBL2. In both cases, the proteins were eluted with three times three CV EB2. The proteins were up concentrated by centrifugation in Amicon Ultra-15 spin concentrators (Merck) with appropriate molecular weight cut offs, in 20 minutes increments at 4°C and 3,200 rcf. The buffers were exchanged to DB2, and the final protein preparations flash frozen in liquid nitrogen and stored at -80°C.

Additionally, both replicase subunits, with tags at either terminus, were also purified in an identical fashion using buffers with 50 mM Tris·HCl (pH 8.0) and 1 mM EDTA instead of 50 mM HEPES (pH7.5) and 10 mM $MgCl_2$, respectively.

2.14 PURE 3.0

2.14.1 Purification of protein components

The protocol for purification and assembly of home-brew PURE 3.0 was adapted form publications by *Shimizu et al.* and *Shepherd et al.* [181,188] and established in the Mutschler group by *K. Libicher*.

For protein purification, small cultures of BL21(DE3) *E. coli* cells, harbouring the expression plasmids pLD1, pLD2, or respectively, pLD3 were grown at 30°C overnight in LB(Kan). The overnight cultures were used to inoculate 1 L LB(Kan) and these fresh cultures were grown at 30°C to an OD_{600} of approximately 0.5. Following induction with 1 mM IPTG, cultures were grown overnight at 16°C. Cells were harvested by centrifugation at 3,200 g and 4°C. For lysis, cell pellets were resuspended in 40 mL/L culture LyB2 and subsequently lysed by sonication on ice. Cell debris was pelleted by centrifugation at 10,000 g and 4°C for 30 minutes. Before being applied to empty gravity flow columns, the supernatant was incubated at 4°C for 30 minutes with 5 mL/L culture of HisPur™ Ni-NTA resin (Thermo Fisher Scientific) equilibrated

with WBL2. Washing and elution steps were performed as followed: Washing five times with five column volumes WBL2, then elution three times with three column volumes EB2. Corresponding elution fractions were pooled and concentrated using Merck Millipore Amicon™ Ultra Centrifugal Filter Units (Thermo Fisher Scientific) with a molecular weight cut-off of 10 kDa for LD1 and LD2, or 3 kDa for LD2, respectively (3200 g at 4°C). After an initial concentration, the buffer was exchanged to DB2 by diluting the protein concentrate in the spin concentrator as much as possible, then up concentrating again. This process was repeated three times in total, and final protein preparations were frozen in liquid nitrogen and stored at -80°C. Tables Table 6 - Table 8 list all individual proteins in the protein fractions LD1 – LD3.

Table 6: Protein components of fraction LD1.

Component	Stock	Reaction
RF1	10 x	1 x
HisRS	10 x	1 x
TyrRS	10 x	1 x
CysRS	10 x	1 x
TrpRS	10 x	1 x
SerRS	10 x	1 x
ValRS	10 x	1 x
MetRS	10 x	1 x
ArgRS	10 x	1 x
GlnRS	10 x	1 x
LeuRS	10 x	1 x
ThrRS	10 x	1 x
LysRS	10 x	1 x

Table 7: Protein components and concentrations of reconstituted fraction LD2.

Component	Stock	Reaction
AsnRS	26 µM	5.2 µM
IleRS	28 µM	5.6 µM
AlaRS	31.25 µM	6.25 µM
PheRS1	50.5 µM	10.1 µM
PheRS2	50.5 µM	10.1 µM
MTF	52.5 µM	10.5 µM
EF-Ts	75 µM	15 µM
IF3	25 µM	5 µM
IF1	84 µM	16.8 µM

Table 8: Protein components of fraction LD3.

Component	Stock	Reaction
AspRS	10 x	1 x
ProRS	10 x	1 x
GlyRS1	10 x	1 x
GlyRS2	10 x	1 x
GluRS	10 x	1 x
RF3	10 x	1 x
RF2	10 x	1 x
RRF	10 x	1 x
EF-G	10 x	1 x
IF2	10 x	1 x

While LD2 purified with this protocol was functional, difficulties with the reproducibility due to plasmid instability led to a change from co-purified LD2 to LD2 reconstituted by mixing individually purified components.

2.14.2 Enzyme mix

Table 9 lists all components of the 6x enzyme mix, as well as their concentrations. Except EF-Tu, all enzymes were purchased from NEB. EF-Tu was prepared as described above.

Table 9: Composition of 6x enzyme mix of PURE 3.0.

Component	Stock	Reaction
T7 RNAP	120 ng/µL	20 ng/µL
Myokinase	30 ng/µL	5 ng/µL
Creatine phosphokinase	60 ng/µL	10 ng/µL
Nucleoside-diphosphate kinase	12 ng/µL	2 ng/µL
RNAse Inhibitor	1.5 U/µL	0.25 U/µL
Inorganic pyrophosphatase	6 U/µL	1 U/µL
EF-Tu	30 µM	5 µM
Ribosomes	10.8 µM	1.8 µM
HEPES·KOH pH 8.0	50 mM	8.3 mM
DTT	6 mM	1 mM
Glycerol	2 %	0.3 %

2.14.3 Energy mix

Table 10 lists all components of the 6x enzyme mix, as well as their concentrations.

Table 10: Composition of 4x energy mix of PURE 3.0.

Component	Stock	Reaction
Potassium Glutamate	400 mM	100 mM
Spermidine	10mM	2.5 mM
ATP	8 mM	2 mM
GTP	8 mM	2 mM
CTP	4 mM	1 mM
UTP	4 mM	1 mM
Sodium Creatine Phosphate	80 mM	20 mM
Folinic Acid	40 mM	10 mM
HEPES-KOH pH 7.5	200 mM	50 mM
Mg(OAc)2	52 mM	13 mM
DTT	20 mM	5 mM
tRNA	OD_{260} 216	OD_{260} 54
Amino Acid mix	4 mM	1 mM

2.14.4 Reaction composition

Per 20 µL reaction, 2 µL of the copurified 10x LD1 and 10x LD3 were mixed with 4 µL reconstituted 5x LD2, then supplemented with 3.33 µL 6x enzyme mix and 5 µL 4x energy mix. Subsequently, the reaction mixes were programmed with the respective RNA templates and further components as described below. Reactions were incubated at 37°C for the designated amount of time.

2.15 Fluorescence assays

2.15.1 Real-time fluorescence assays

All fluorescence-based readout constructs are based on designs by Weise *et al.* [180]. The experiments using the PURExpress *In Vitro* Protein Synthesis Kit (NEB) were prepared in standard 25 µL reactions, supplemented with DFHBI-1T, MS2rep, $[rep]_{MSRP}$ and $[F30\text{-}Bro(-)]_{UTRs(-)}$ or $[F30\text{-}Bro(+)]_{MSRP(-)\text{-}1.0}$, respectively, as described for the individual experiments:

Transcription from $[F30\text{-}Bro(-)]_{UTRs(-)}$ *in PURExpress*: 300 nM MS2rep, 150 nM $[F30\text{-}Bro(-)]_{UTRs(-)}$ and 10 µM DFHBI-1T

IVTxT of MSRP-22 variants in PURExpress: 100 nM $[rep]_{MSRP}$ or RZ, 100 nM $[F30\text{-}Bro(-)]_{MSRP(-)\text{-}1.0}$ and 10 µM DFHBI-1T

All real-time fluorescence experiments based on the homemade PURE system (PURE 3.0 / PUREred) were assembled including 0.5 mM ATP, GTP, CTP and UTP, 10 mM DTT, 15 µM EF-Tu, 1.5 µM S1, DFHBI-1T and all other components as followed, as 20 µL reactions.

Transcription from $[F30\text{-}Bro(-)]_{UTRs(-)}$ *in PURE 3.0*: 300 nM MS2rep, 150 nM $[F30\text{-}Bro(-)]_{UTRs(-)}$, 1x LD1, reconstituted LD2 and LD3, and 10 µM DFHBI-1T

LD2 depletion experiments: 300 nM MS2rep, 50 nM $[F30\text{-}Bro(-)]_{UTRs(-)}$, 5 µM LD2 components, and 10 µM DFHBI-1T

Single LD2 component: 300 nM MS2rep, 50 nM $[F30\text{-}Bro(-)]_{UTRs(-)}$, 15 µM EF-Ts, 5 µM LD2 components, and 10 µM DFHBI-1T

Co-factor titration: 300 nM MS2rep, 50 nM $[F30\text{-}Bro(-)]_{UTRs(-)}$, 15 µM EF-Ts, and 10 µM DFHBI-1T

MS2-Bro replication: 1 µM nM MS2rep, 50 nM $[F30\text{-}Bro(-)]_{UTRs(-)}$, 15 µM EF-Ts, 10 µM DFHBI-1T, and 0.5 U per µL RNase inhibitor (moloX)

Untemplated replication: 200 µM nM MS2rep or Qβrep, 15 µM EF-Ts, and 1x SYBR™ Green II Nucleic Acid gel stain (Thermo Fisher Scientific)

IF3 inhibition: 300 nM MS2rep, $[F30\text{-}Bro(-)]_{UTRs(-)}$ (as indicated), 15 µM EF-Ts, IF3 (as indicated), and 10 µM DFHBI-1T. Reactions were overlaid with 20 µL silicon oil (PDMS, viscosity 20 cSt, Sigma-Aldrich), fluorescence was measured with 12.5%FAM-LED intensity.

The final concentration of $MgCl_2$ was 6 mM, with HEPES, KCl and glycerol supplemented in 50 mM, 100 mM and 18 %, respectively. All reactions were prepared in MicroAmp Fast 8-TubeStrips (Thermo Fisher Scientific) and, except for IF3 inhibition kinetic experiments, incubated at 37 °C in a StepOne Real-Time PCR System (Thermo Fisher Scientific). The fluorescence signals were recorded every minute over four hours or every two minutes for six hours for experiments with $[F30\text{-}Bro]_{UTRs}$ and $[F30\text{-}Bro]_{MSRP(-)\text{-}1.0}$, or $[F30\text{-}Bro]_{MS2}$, respectively. IF3 inhibition kinetic measurements were performed in a T-16 ISO fluorescence system (AXXIN), with the time between measurements set to the possible minimum (approximately 30 seconds).

2.15.2 Fluorescence anisotropy

For the fluorescence anisotropy assay, $[F30\text{-}Bro(+)]_{UTR(-)}$ was added in a final concentration of 200 nM to a minimal buffer mix, without MS2rep subunit, EF-Tu, EF-Ts and S1, supplemented with 150 nM DFHBI-1T. After addition of either IF1 or IF3, the reaction mix was incubated at room temperature for 15 minutes and fluorescence polarization was measured three times per replicate. The anisotropy was subsequently calculated using following relation between polarization (P) and anisotropy (A):

$$A = \frac{2 \cdot P}{3 - P}$$

Fluorescence anisotropy was measured in a Clariostar^{Plus} with the following fluorescence settings: λ_{exc} = 482 nm (bandwidth +/- 8 nm) and λ_{em} = 530 nM (bandwidth +/- 20 nm) with dichroic longpass filters (504 nM) in black 384-well plates with clear bottom (Greiner) at 25°C.

2.16 Electrophoretic mobility shift assay

EMSAs performed to detect interaction between IF1 or IF3, respectively, with $[F30\text{-}Bro(-)]_{UTR(-)}$ were done by mixing all components in the minimal buffer system for replication, minus the other co-factors. Additionally, an appropriate volume of 6.7x EMSA B1 was added to a final concentration of 1x. After incubating this mix at room temperature for 15 minutes, samples were loaded onto agarose gels and the gels were run as described in section 2.3.1.
In the case of probing for interactions between the MP and genomic MS2, or respectively, MSRP22 derivatives, the samples were prepared the same way. However, buffers PBB and EMSA B2 were used for the sample preparation.

2.17 *In vitro* replication assays

2.17.1 Serial transfer and sequencing of MSRP-22

The reactions for the serial transfer experiment were prepared as the samples for real-time fluorescence assays, with 50 nM of genomic MS2 RNA, 1 µM MS2rep, 15 µM IF1, 15 µM IF3 as indicated but without DFHBI-1T. All reactions were incubated at 37°C for 3 hours. Of these reactions, 4 µL were used to program the following round of 20 µL reactions. Samples for gel electrophoresis were prepared by adding one volume of 2x RNA loading dye to the reaction mix and incubating the samples at 70°C for 5 minutes.
To obtain sequencing data of the emerging small RNAs, the RNAs were purified from the serial transfer reactions. Polyadenylation of these RNAs with *E. coli* Poly(A) Polymerase (NEB) was followed by another RNA clean-up. The first strand cDNA was synthesised using SuperScript IV First-Strand Synthesis System (Thermo Fisher Scientific) in with the primer CDSII-T_{24}VN, subsequently purified, and polyadenylated with dATP and Terminal Transferase (NEB). After

RNA digest with RNAse H (NEB), the single stranded RNA was purified, and the second strand synthesised in a 10-cycle PCR using Q5 (NEB) with the primers CDSII-T_{24}VN and CDSII (0.1 and 1 µM). Finally, the purified double-stranded cDNA was inserted into a cloning vector using the NEB® PCR Cloning Kit (NEB). Finally, 2 µL of the reactions were transformed into chemically competent Top10 *E. coli* cells. Clones from this transformation were used for plasmid isolation, followed by Sanger sequencing.

The sequence identity of RNAs amplified in the initial replication assays was confirmed by reverse transcribing purified, polyadenylated RNAs. Using the 2nd Strand cDNA Synthesis protocol for the Template Switching RT Enzyme Mix (NEB) and primers CDSII-T_{24}VN and TSO-CDSII, double-stranded cDNAs were synthesized from the polyadenylated RNAs. The cDNAs were isolated and inserted into a cloning vector using the NEB® PCR Cloning Kit (NEB). Subsequently, 2 µL of the reactions were transformed into chemically competent Top10 *E. coli* cells. Clones from this transformation were used for plasmid isolation, followed by Sanger sequencing.

2.17.2 Cloning of MSRP-22 based replicators

The plasmids encoding for the [F30-Bro]$_{MSRP}$ constructs (pUCIDT_MSRP-1.0(F30-Bro), pUCIDT_MSRP-2.0m(F30-Bro), pUCIDT_MSRP-2.0p(F30-Bro), pUCIDT_MSRP-3.0m(F30-Bro), pUCIDT_MSRP-3.0p(F30-Bro)) were ordered from IDT. The plasmids encoding for the corresponding [*zeoR*]$_{MSRP}$ constructs (pUCIDT_MSRP-1.0(zeoR), pUCIDT_MSRP-2.0m(zeoR), pUCIDT_MSRP-2.0p(zeoR), pUCIDT_MSRP-3.0m(zeoR), pUCIDT_MSRP-3.0p(zeoR)), as well as all plasmids encoding for [*rep*]$_{MSRP}$ constructs (pUCIDT_MSRP-1.0(rep), pUCIDT_MSRP-2.0m(rep), pUCIDT_MSRP-2.0p(rep), pUCIDT_MSRP-3.0m(rep), pUCIDT_MSRP-3.0p(rep)) were assembled by HiFi assembly of PCR amplicons of the respective coding sequences and vector backbones. Plasmids encoding for the *rep* and *zeoR* encoding constructs (pUCIDT_RZ-1, pUCIDT_RZ-2, pUCIDT_RZ-3, pUCIDT_RZ-4, pUCIDT_RZ-5) were assembled in a similar way, however, here the *zeoR* encoding insert was integrated into the respective *rep* encoding vector backbone.

2.17.3 Replication of MSRP-22 derived constructs

The reactions for the detection of *in vitro* replication of MSRP-22 based RNAs were prepared as described above, with 100 nM of the RNA template, 250 nM MS2rep. All reactions for one specific template were prepared as a single master mix, then split into the number of individual reactions, according to the time points taken, and incubated at 37°C. After the incubation, samples were mixed with an equal volume of 2x RNA loading dye, heated up to 70°C for 5 minutes and analysed by gel electrophoresis.

2.18 *In vivo* RNA replication assays

For the trial assays to detect *in vivo* replication of [*zeoR*]$_{MSRP}$ constructs, *E. coli* F$^+$5695 cells were first transformed with the plasmids pBAD33_His$_6$-MS2rep and pBAD33_S1-His$_6$, respectively. Correct transformants were identified through sequencing and used to inoculate starter cultures. Cultures were grown at 30°C overnight, diluted to an OD_{600} of 0.05 and grown for one hour at 30°C. After induction with 0.2% L-Arabinose, cells were grown for 90 minutes at 30°C, made electrocompetent and subsequently electroporated with 2.5 µg of the respective RNAs, water, or respectively, the corresponding plasmids (50 ng). Cells were allowed to recover in LB at 37°C, shaking at 900 rpm, for 5 / 15 / 30 / 60 minutes and plated on 1.5% LB agar plates containing 12.5 µg/mL Zeocin and 0.1% L-Arabinose. The plates were incubated at 37°C overnight. For the detection of *in vivo* replication of RZ-1, RZ-2, RZ-3, RZ-4 and RZ-5, respectively, *E. coli* F$^+$5695 cells were used. Here, cells were treated identically, however, no L-Arabinose was added to the liquid media.

For the quantification of the *in vivo* effect of the RZ RNAs, or respectively, the respective plasmids (25 ng for RZ and [*zeoR*]$_{MSRP-1.0}$), cells were treated as described above. However, cells were plated on plates containing 10 µg/mL Zeocin in case of RNA transformation, or respectively, 1/10th (100 µL) of the transformed cells were plated on plates containing 37.5 µg/mL in case of plasmid transformation.

Colony PCRs were performed by using GoTaq G2 Hot Start Master Mix (Promega) with primers colPCR_1 and colPCR_2 (both 200 nM) and running the following cycling program:

95°C	60 seconds	
95°C	10 seconds	¯|
60°C	30 seconds	| 40 cycles
68°C	95 seconds	_|
68°C	10 minutes	
15°C	hold	

2.19 Reproducibility

In general, if not stated otherwise, all quantitative experiments were repeated as triplicates, and all proofs of concept as a single experiment. For all *in vitro* experiments, the replicates were performed as technically independent replicates using the same enzyme stocks but were always separately prepared and executed. Replicates of *in vivo* experiments were always biologically independent replicates, using the same strains and enzyme stocks, but always individually prepared starting cultures and master mixes.

3. Results

3.1 Mutagenic IVT of RNA hairpins

The MS2 maturation protein is the key component of the RNA delivery to *E. coli* harbouring the fertility plasmid, as it directly links the cargo-RNA to the host cell (Figure 6A)[107]. However, to establish the MP-RNA interaction *in vitro* with artificial RNA genomes, the required RNA motifs first need to be identified. Consequentially, systemic evolution of ligands by exponential enrichment (SELEX) was the best approach for this (Figure 6B). Instead of using a completely randomized library for the SELEX, though, hairpins that are closely located to the MP within the MS2 virion were used as a starting point to synthesize the initial pool of mutated sequences (Figure 6C) [105]. The general idea behind this was that the mutants still adapt similar secondary structures as the wildtype sequence. This sequence binds to the MP without disrupting the ensuing infection. Therefore, the mutant sequences should ideally also have a preference to bind the correct epitope, albeit with higher affinity.

To generate this initial pool, an IVT template was designed, consisting of a T7 promoter, followed by the two RNA hairpins between nucleotides 1748 and 1791 in the MS2 genome and connected to the hairpin from nucleotide 3540 to 3564, via a linker of 7 (CH7, CACACAC) or respectively, 9 (CH9, CACACACAC) nucleotides length. In the following mutagenic IVT, biased NTP pools and manganese (II) ions were used to decrease the fidelity of the T7-RNAP [192] and thereby produce a pool of sequence diverse mutants of the combined hairpin module. To gain insight into the mutation rate under the mutagenic conditions used for the first IVT, the same hairpin modules were also *in vitro* transcribed under regular, non-mutagenic conditions. Samples of these RNA pools were handed over to the core facility of the MPI of Biochemistry (Martinsried, Germany) for synthesis of the corresponding cDNAs and subsequent Illumina next generation deep sequencing.

By comparison of the relative base read counts per position in the RNA between the mutated and the control pool, an estimate of the mutagenesis rate was possible (Figure 7A, B). Regarding the overall distribution of mutations, the observed pattern for CH7 and CH9 are highly similar, although the absolute mutation rates per position are different. While the more extreme, both low and high rates, appear towards the 5’-end of the sequence, the extremes even out towards the 3’-end to more stable mutation rates. The highest observed mutation rates occurred at positions 5 and 18, both occupied by A in the reference sequence. A hypothesis as to why the mutations are distributed in this pattern might be, that especially at the beginning of the transcription, misincorporation of nucleotides might abort this process more readily. In addition, G and C nucleotides potentially stabilise the T7-RNAP/DNA/RNA-complex more than A or T/U, as more hydrogen bonds form during base pairing (G-C: 3, A-T/U: 2).

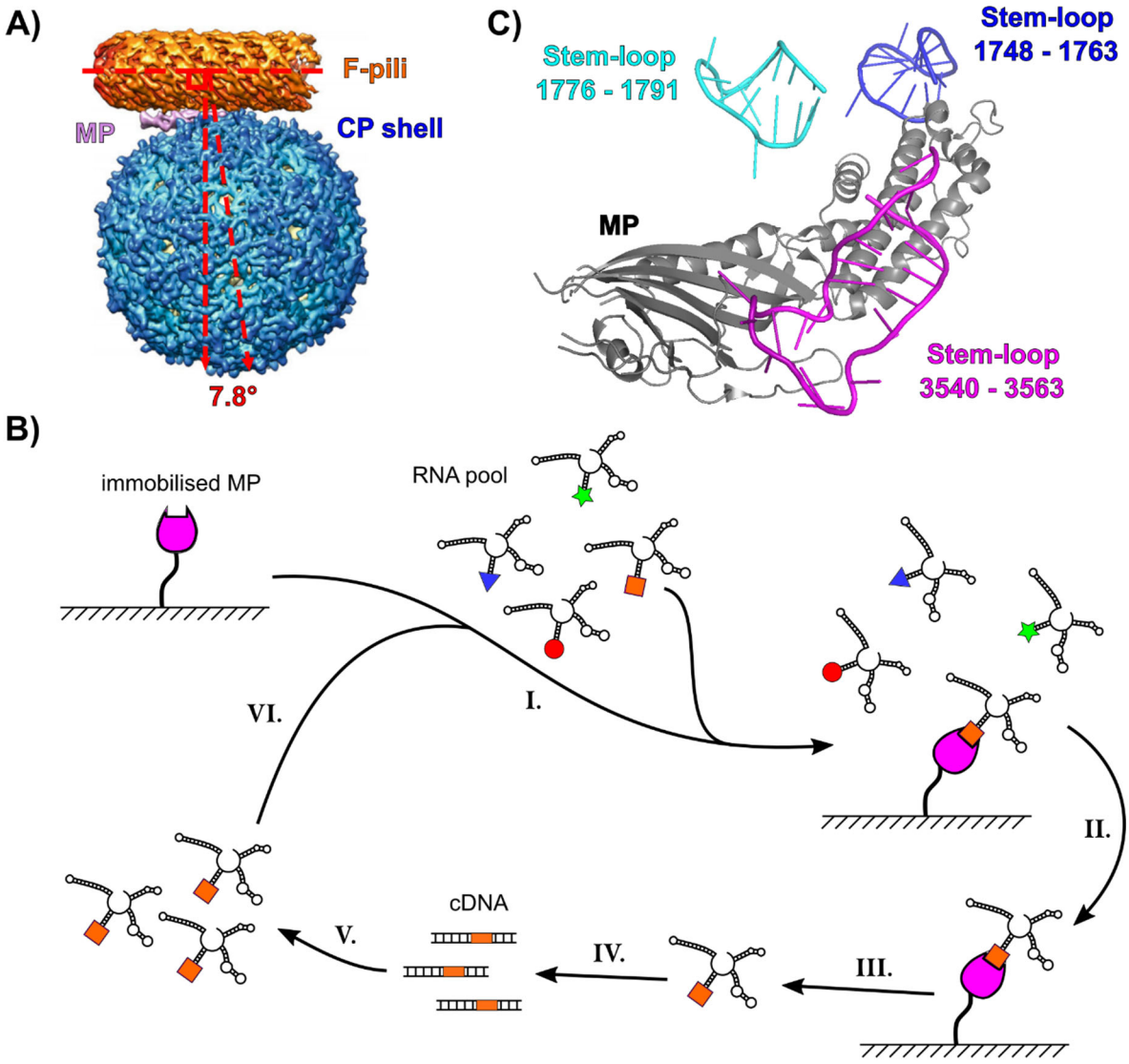


Figure 6 Fundamentals of maturation-protein-based RNA delivery: A) MS2 capsid (blue) bound to a F-pili (orange). The MP (pink) is at the interface between pili and capsid, tilting the two-fold axis of the capsid by 7.8°C (red lines). The image was adapted from published cryo-EM data from *Meng et al.* [107]. **B)** Simplified scheme of SELEX for the identification of MP-binding RNA motifs. (**I.**) A pool of RNAs with randomised sequences is added to immobilised MP. RNAs that adopt a suited structure can bind the MP. (**II.**) Unbound RNAs are washed out, while bound RNAs are retained by their interaction with the MP. (**III.**) Bound RNAs are eluted from the MP and (**IV.**) subsequently reverse transcribed into the corresponding complementary DNA (cDNA) and amplified by PCR. (**V.**) The double stranded DNA (dsDNA) is used for *in vitro* transcription of a new pool of RNAs. As only binding RNAs are carried over to the previous steps, these sequences are enriched in the new pool. (**VI.**) The new pool can consequentially be used for a new round of selection. After several iterations of this cycle, only the sequences with the strongest binding interaction will be left in the pool. **C)** Cryo-EM structure of MP (grey) and proximal RNA stem loops (magenta / light blue / dark blue). Numbers refer to the nucleotide position in the genomic MS2 RNA. The image was prepared using open source PyMOL version 2.6.0a0 [208]. Data was derived from cryo-EM data published by *Dai et al.* (PDB: 5TC1) [105].

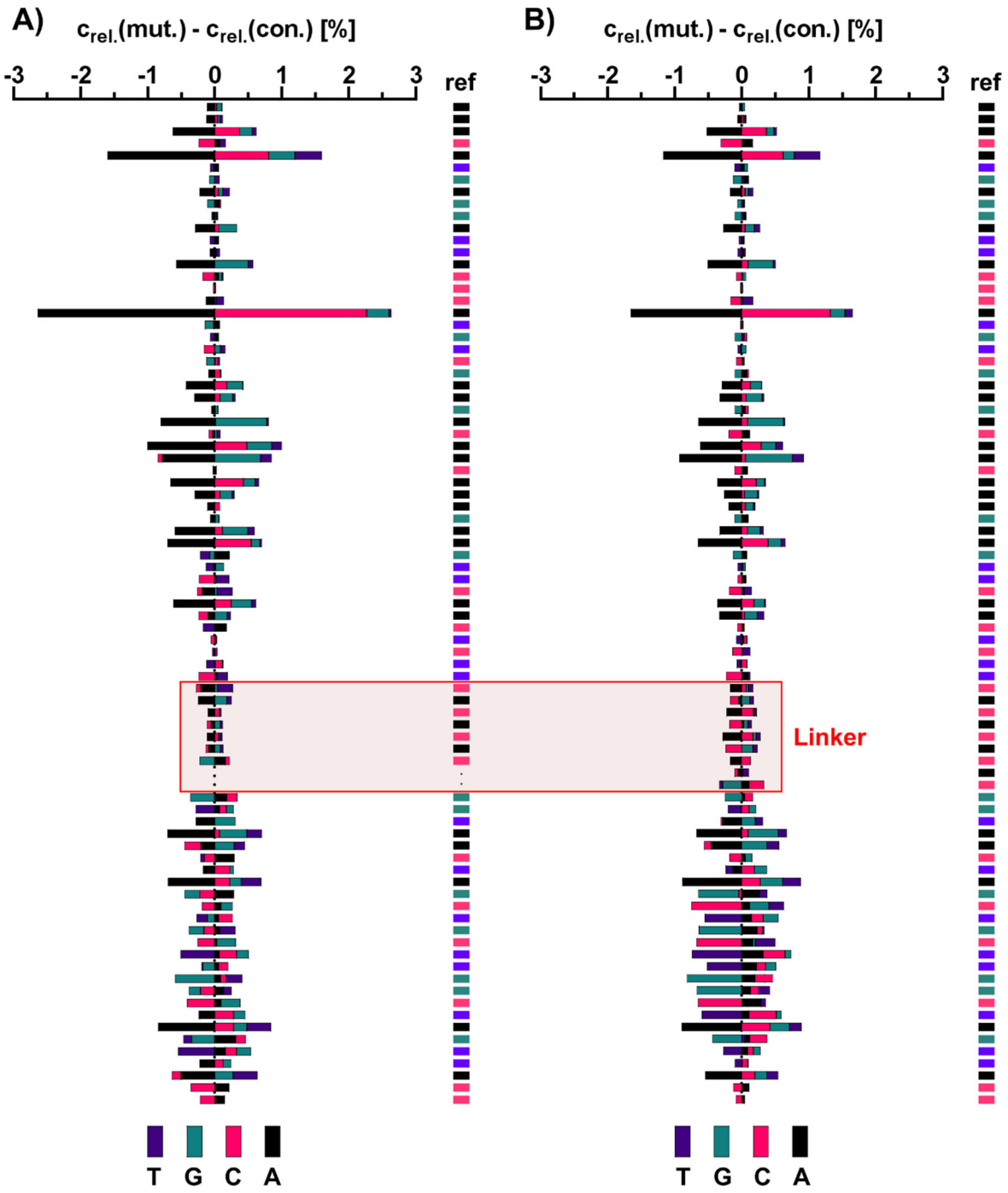


Figure 7 Mutation rates for mutagenic *in vitro* transcription: A) and **B)** Difference of relative base read counts (c_{rel}) between cDNAs of combined hairpin RNAs from either mutagenic (mut) or a regular control IVT (con.). The differences between mutated and control pool are depicted for each base and position in the combined hairpin RNA with a linker length of 7 nucleotides (A)) and 9 nucleotides (B)), respectively. The bases are coded by colour: A = black, C = magenta, G = petrol, T = purple. The red box highlights the linker region combining the two RNA hairpin modules. The coloured columns depict bases of the reference sequence (ref). Values below 0 indicate that the respective base occurs with lower frequency as in the control pool, values above 0 indicate that the respective base occurs with higher frequency. The sequencing data was derived from Illumina next generation deep sequencing.

Thus, if either G or C are replaced with another nucleotide, abortion could more frequently occur during early transcription, before the T7-RNAP can switch into a more stable, processive

state.

In general, positions that were occupied by As appeared to be more prone to misincorporation of nucleotides than other positions, due to a seemingly higher overall mutation rate. Though, the individual nucleotides do not make up equal parts of the sequence, with A (~35 %) being the most abundant, followed by C (~25 %), T (~22 %) and G (~18 %). Therefore, the higher rates for mutations of positions with As might be partially due to overrepresentation. The average mutation rate per position, defined as the sum of all mutation rates per base per position above 0 divided by the total number of positions, was calculated to be 0.355 % for CH7 and 0.338 % for CH9, respectively. Comparing these rates to rates reported for error prone PCR (0.33 – 3.3 [209], 0.66 % [210]), this method for the generation of mutant libraries appears comparably efficient. This method might be especially useful for selection experiments when combined with a preceding error prone PCR.

3.2 Detecting RNA transfection

The idea of RNA delivery with the MP, led to the question of how to detect successful RNA delivery. While plaque assays pose a robust way to detect delivery of MS2 RNA, it was not applicable to the delivery of non-infectious and non-self-replicating RNAs. This is due to the sensitivity of RNAs to degradation and hydrolysis, as well as the lack of any means to persist through cell divisions at a stable level. A potential system to overcome the limited persistence of non-replicating RNAs cells was the group IIB intron EcI5.

This type of group II introns consists of the catalytically active intron RNA and an intron encoded protein (IEP), (Figure 8A) [195,197,211–213]. In a first step, the intron gets transcribed into RNA, flanked by the co-transcribed parts of the 5'- and 3'-exon. After this, the IEP can be translated, bind to the intron, and assist the RNA in self-splicing, resulting in the intron forming a lariat ring, while the exon fragments are ligated to each other [214]. The lariat-IEP RNP then scans the DNA for a potential insertion site, mostly determined by sequence complementarity between the DNA and intron RNA. Subsequently, the intron reverse splices itself into the target DNA strand, while the IEP cleaves the second DNA strand, generating a free 3'-OH. Using this new 3'-end as a primer, the IEP starts synthesizing cDNA by reverse transcribing the inserted intron RNA. Finally, removal of the RNA, and DNA repair generate a new locus encoding for the group II intron, from which the cycle could restart [211–213].

There are two characteristics of these systems, that make them especially interesting for this study. Firstly, with exception to a very limited number of nucleotides, the insertion site can almost freely be programmed by changing specific nucleotides in the intron RNA (Figure 8B) [195,197]. Secondly, the activity of the RNA is not lost after partial removal of the ORF, but potentially increases, if the IEP can be provided in trans [195].

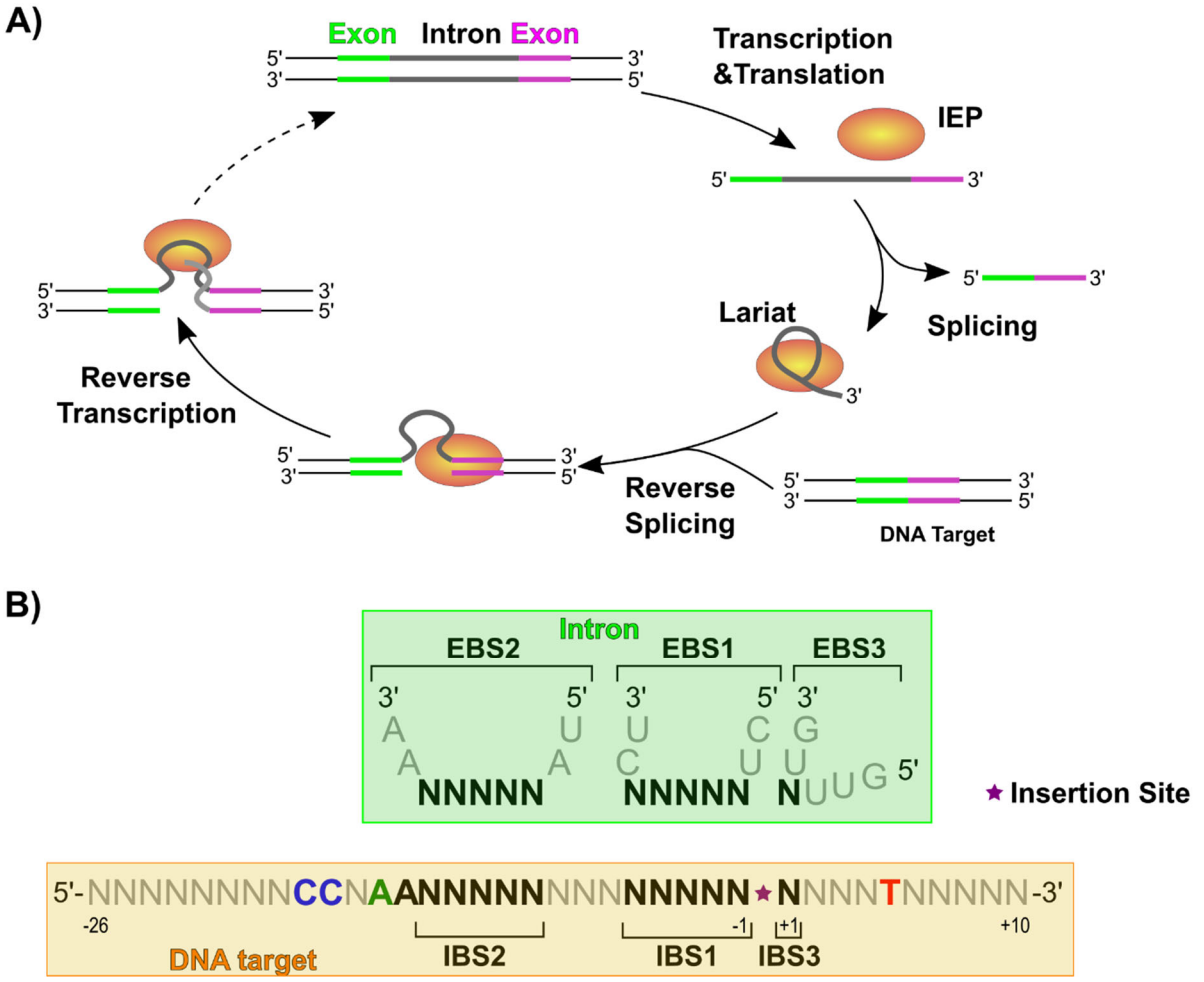


Figure 8 Group II intron EcI5: A) Mechanism of intron mobility: Following transcription of the intron RNA, embedded within two exons, the intron encoded protein (IEP) can get translated. Binding of the IEP to the intron RNA facilitates self-splicing of the intron out of the exons, thereby forming the RNA lariat structure. The RNP subsequently scans DNA for a suitable insertion site, comprised of binding sites for the IEP, as well as recognition sequences for base pairing with the intron RNA itself. Catalytic activity of the intron RNA leads to reverse splicing of the intron into target DNA strand, followed by cleavage of the second DNA strand by the IEP. Subsequently, reverse transcription of the intron RNA by the IEP, using the 3'-end of the second DNA strand as a primer, ends in the synthesis of cDNA. Finally, removal of the RNA and DNA repair produce a new locus of DNA harbouring the intron sequence [195,197,211–214]. **B)** Detailed representation of the important elements for EcI5-intron/DNA interaction, reverse splicing, and reprogramming both on the RNA (green box) and the DNA (orange box). Positions -26 to -14, as well as +2 to +10 serve as binding sites for the IEP. While most of these bases can be variable (grey, regular Ns), positions -18 and -17 (blue, bold Cs), -15 (bold, green A), as well as +5 (red, bold T) are highly invariable. Noteworthy, although the A at -14 (bold, black) was seemingly essential for splicing activity in the selection experiment by *Zhuang et al.* [195], if the corresponding base in the intron is changed accordingly, any nucleotide can be chosen for this position [215]. The intron binding sites (black, bold Ns) on the DNA at positions -13 to -9 (IBS2), -5 to -1 (IBS1) and +1 (IBS3) can be freely chosen, however they must be complementary to the corresponding exon binding sites EBS 1 – 3 on the intron RNA (black, bold, Ns), and vice versa. The intron inserts itself between IBS1 and IBS3 (purple star) [195].

In a first step, an already described version of the EcI5 system, targeting the *lacZ* gene of *E. coli* (EcI5(ΔORF)-lacZ-1806s), was used in a blue/white screening assay [195]. The results of this assay revealed that the EcI5 variant does indeed efficiently disrupt the *lacZ* gene of the *E. coli* strain HMS-174(DE3), even in the absence of transcription inducing IPTG (Figure 9, Supplementary Figure 2). Remarkably, even without preceeding induction, no blue colonies were detected for this intron. In contrast, an intron that did not target *lacZ* (EcI5(ΔORF)-sacB-491s) did not disrupt the *lacZ* gene. This implies that the insertion already efficiently takes place with the background expression from a leaky T7 promoter, and that the insertion is sequence specific. The induction with IPTG drastically reduced the overall number of colony forming units (cfu). This presumably was caused by a combination of toxicity of IPTG itself, high metabolic burden and stress due to overexpression and elevated activity of the intron and ther IEP, as well as potential off-target insertions [216].

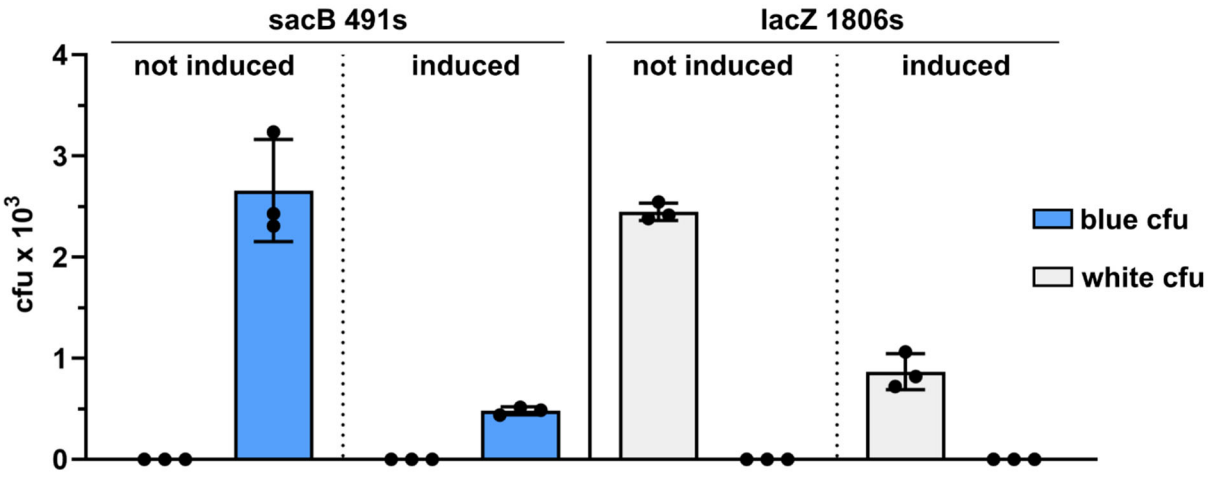


Figure 9 Disruption of *lacZ*: Blue/white screening of HMS-174(DE3) cells that either harboured a plasmid encoding the EcI5(ΔORF) construct targeting the sense strand (s) of the coding sequence of the *sacB* gene at position 491 (sacB 491s), or respectively, the *lacZ gene* at position 1806 (lacZ 1806s). Hereby, cultures were induced with IPTG for the transcription of the EcI5 intron and then subjected to blue/white screening. Blue and white cfu were counted after incubation at 37°C for 24 hours. Non induced cultures were used as additional controls. Error bars indicate the standard deviation over three technical replicates, using cultures grown from single colonies with confirmed sequence identity of the respective plasmids.

Using blue/white screening to identify clones with genes disrupted by the EcI5 intron is possible. However, when the intron and the IEP can not be readily transcribed and translated at a sufficient level, as might be the case for RNA delivery of the intron/IEP-mRNA, the percentage of white colonies within the total population might be drastically lower. Hence, identification of clones harbouring a disrupted *lacZ* gene might become more difficult, or even impossible. The alternative was to do a negative selection instead, in this case using the *sacB* gene from *Bacillus subtilis* in combination with sucrose. The *sacB* gene encodes for the

enzyme levansucrase, which catalyses the polymerisation of sucrose into levan. Levan is a high molecular weight polysaccharide and toxic for several gram-negative bacteria, including *E. coli* [217–219]. As only colonies with disrupted *sacB* gene would be able to survive the selection with sucrose, the detection of successful RNA transformation would be substantially facilitated.

To integrate the *sacB* gene into the genome of *E. coli* strain HMS-174(DE3), the Red/ET system was chosen, which allows precise genomic insertions of a linear gene cassette, based on homologous recombination (Figure 10) [220–222]. Thereby, a linear cassette, encoding for a selection marker, additional genes of interest and with 5'- and 3'- ends homologous to a target genomic locus, gets transformed into cells that harbour a plasmid for the expression of the Red/ET recombination system.

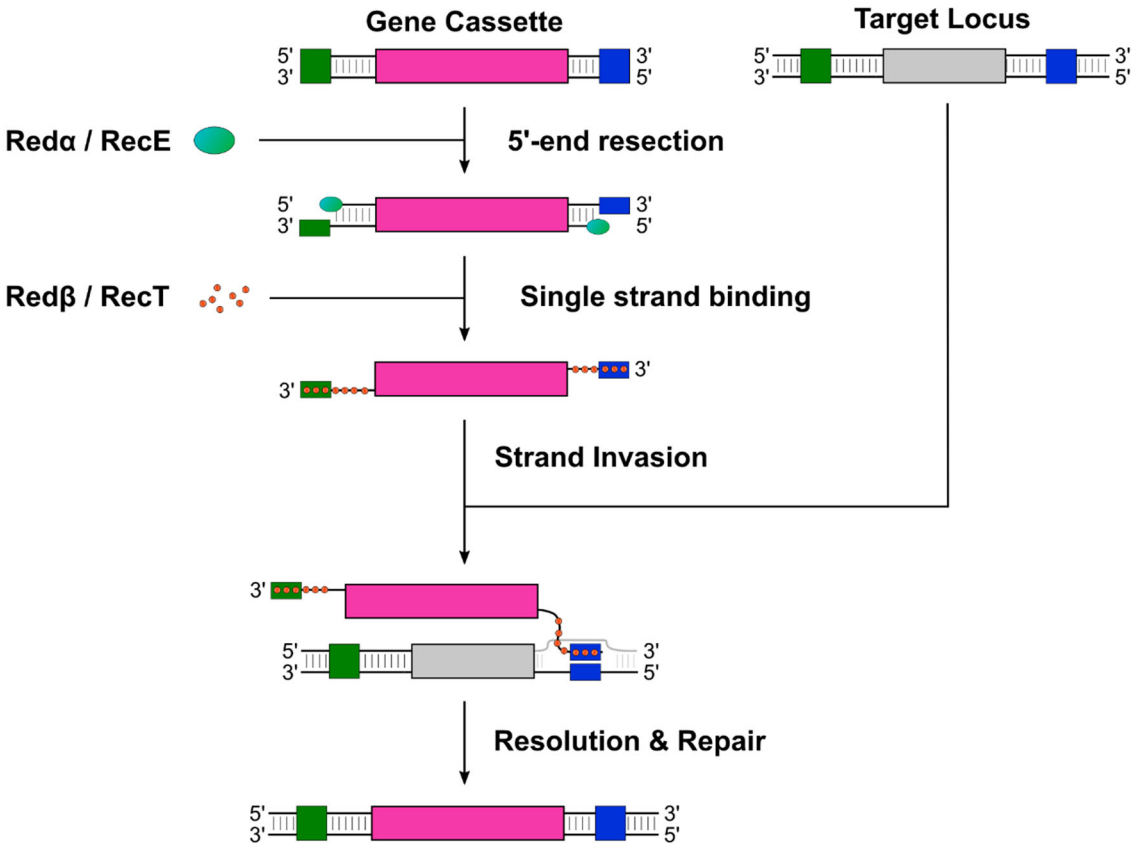


Figure 10 The RecET system and sacB disruption with Ecl5: Mechanism of homology driven recombination mediated by the rac prophage RecE/RecT system, or respectively, λ phage Redα/Redβ system. Resection of the linear gene cassette, harbouring a selection marker, as well as further genes of interest, by exonucleases RecE or Redα produces 3'-overhangs. These overhangs are recognized by the corresponding annealing proteins RecT or Redβ, respectively, protecting the ssDNA strands and the cassette in total from degradation. In addition, the annealing proteins enable repair/recombination via single strand joining (not shown) if a homologous donor is provided. Alternatively, strand invasion of the single stranded ends of the cassette into the double stranded homologous region of the target locus can occur. For simplicity, the strand invasion is only shown for one end. Following the resolution of the intermediary formed D-loop and DNA repair, a new locus is generated where the targeted sequence was replaced with the cargo gene [222,223].

Following the transformation, the linear cassette gets processed by the exonucleases RecE or, respectively Redα, leaving a single stranded overhang at the 3'-end. This overhang subsequently gets bound by RecT (for RecE) or Redβ (for Redα), which prevents degradation of the linear gene cassette. After homology driven strand invasion of one overhang into the respective homologous stretch of the target locus, an intermediary D-loop structure forms. This process occurs in parallel on the other end of the cassette. Resolution of the D-loop and subsequent DNA repair complete the replacement of the target sequence with the gene cassette [222,223].The cassette and Red/ET system used in this study are derived from the commercial Quick & Easy *E. coli* Gene Deletion Kit (Gene Bridges). For the genomic integration, the gene cassette was expanded with the constitutive P70a promoter [194], followed by the sacB gene (Supplementary Figure 1). The *rnc* gene, which encodes for the dsRNA specific RNAse III, was chosen as target locus [224]. Preliminary tests with *E. coli* HMS-174(DE3)^{sacB} cells showed that integration of the *sacB* gene was indeed successful and also conferred lethal sensitivity to sucrose starting at concentrations as low as 2.5 % (data not shown).

Based on the publications by *Zhuang et al.* [195] and *Perutka et al.* [197], an algorithm was written (see Section 2.6.2 & 7.5) that allowed the identification of potential target sites for EcI5 insertion. Thereby, potential sequences are identified based on a score, calculated for 36 bp long increments of the input sequence. The individual scores themselves are computed from nucleotide frequencies observed in a selection experiment by *Zhuang et al.* [195]. Furthermore, *Zhuang et al* identified several potential positions in the sense strand (s) of the *lacZ* gene, including the top three hits at position 163, 1806 and 2427, with calculated scores of 11.72, 11.01 and 11.18, respectively [195]. For the used parameters, the score could in theory range between approximately -29 and +21. In an ensuing *in vivo* experiment, only sites with scores above 8.2 could be efficiently disrupted [195]. Thus, this value presumably represents the cut off threshhold for the prediction of a potential insertion site. Noteworthy though, the observed disruption efficiencies, based on blue/white screening, did not necessarily correlate with the calculated scores [195].

Using the algorithm written for this thesis with similar parameters as *Zhuang et al*, these three top hits were found in the *lacZ* gene of *E. coli* strain K12 (NCBI Gene ID 945006) [195]. The calculated scores were 11.82 (163s), 10.8 (1806s) and 10.4 (2427s). This aligns well with the reported values, considering potential deviations due to rounding and potential differences in the input sequences.

The subsequent search for suited insertion sites within the *sacB* gene (NCBI Gene ID 936413) resulted in three hits with a score above 7.5, a threshold set in accordance with the results of *Zhuang et al.* [195]. These three sites were 116s (8.19), 309s (8.92) and 491s (7.81). According

to these predictions, the corresponding plasmids were cloned and tested for their potential to disrupt the genomic *sacB* locus of *E. coli* HMS-174(DE3)^{sacB} (Figure 11).

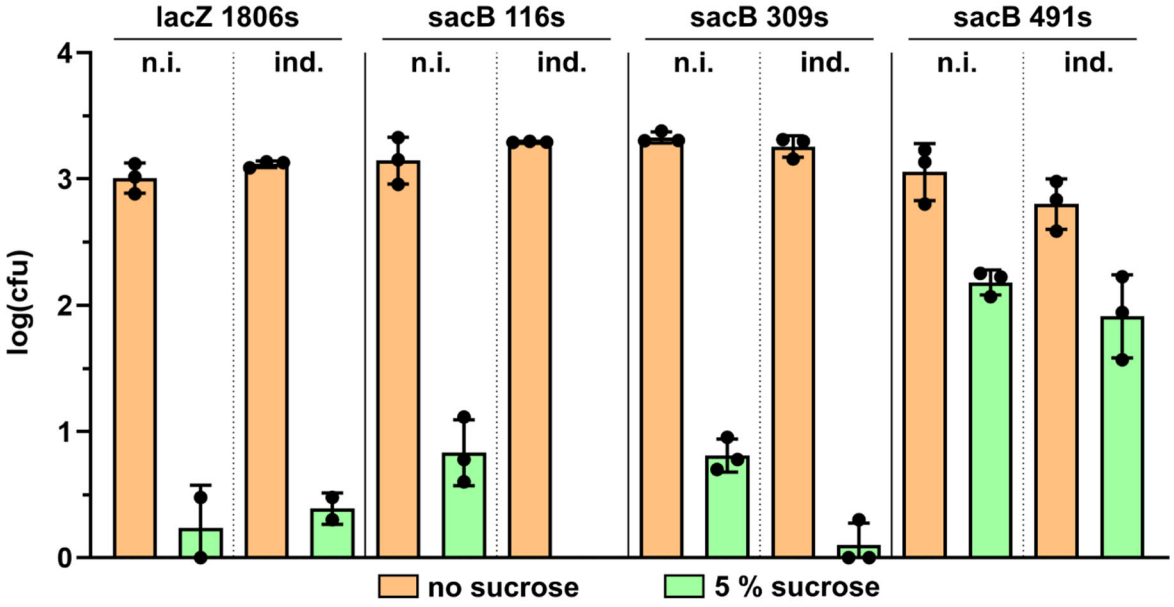


Figure 11 Disruption of *sacB*: Observed cfu for the sucrose selection of *E. coli* HMS-174(DE3)^{sacB} cells harbouring plasmids encoding for one of the three EcI5 introns targeting *sacB* (sacB 116s / sacB 309s / sacB 491s), or respectively, *lacZ* (lacZ 1806s). Cultures were induced (ind) with 0.1 mM IPTG, grown at 37°C for three hours and then plated on plates supplemented with (green) or without (orange) 5 % sucrose. The intron targeting *lacZ*, as well as the non induced cultures (n.i.), served as negative controls. Error bars show the standard deviation of the mean, based on three technical replicates. Error bars indicate the standard deviation over three technical replicates, using cultures grown from single colonies with confirmed sequence identity of the respective plasmids.

As negative control, the EcI5 intron targeting *lacZ* (lacZ 1806s) was used. As expected, the presence of sucrose almost completely inhibited cell growth when the *sacB* gene was not targeted. However, even when *sacB* was targeted at positions 116s and 309s, cell growth did not substantially, if at all, exceed the background level. In contrast, position 491s could be successfully targeted for gene disruption, albeit the number of cfus was reduced by one magnitude compared to conditions without sucrose. The scores calculated for the target sites for *sacB* (8.19 for 116s, 8.92 for 309s, 7.81 for 491s) were lower than for position 1806s in the *lacZ* gene (10.8). Thus, the results seemingly reflect the expected lower insertion and disruption efficiency, based on the calculated scores. However, the following effects, that were not considered for the score calculation, might have contributed to lower disruption rates.

First, it was shown that some group II introns can also splice into ssRNA *in vitro*, if that RNA contains the respective target sequences and this sequence is accessible [225]. As this could also occur *in vivo*, the intron lariat formed from the overexpressed precursor potentially targets both its target DNA, as well as the mRNA transcribed from the target locus, thereby disrupting the gene on both levels. On the other hand, the insertion into the target locus allows

transcription of new intron RNA, embedded into the respective mRNA. As the intron can splice out and thereby ligate the exons, the wildtype mRNA could be restored, and the encoded protein translated [214]. These two opposing splicing activities are presumably both modulated by the IEP. In addition, they might also be influenced by the secondary structures of the flanking exons, based on if they promote or disturb proper folding, or respectively, block the insertion of the intron. Thus, the observed disruption rate would also depend on the frequency and ratio of these splicing events, and how they affect the levels of functional mRNA. Importantly, also the position of the insertion with respect to the catalytically relevant epitopes of the protein impacts the read-out. Here, if the rescue of the levansucrase-mRNA would be favoured, or the catalytic center would still be intact, the levels of active levansucrase could be sufficient to confer a substantial sensitivity to sucrose. In consequence, less colonies would grow. The result would be false negative, even if insertion into the *sacB* locus took place.

Another factor could be the lifetime and expression level of the levansucrase inside the cell. If the half-life of the enzyme is sufficiently long, and cellular concetration and/or activity are high enough, the incubation time after induction and post intron insertion could be too short. In consequence, even if the gene would be disrupted, the residual enzyme activity might still suffice to result in lethal concentrations of levansucrase. This could explain the overall lower efficiency, if the sensitivity for the *sacB* based readout is higher than for the blue/white screening.

However, as the induction did not, within the range of error, noticeably increase the number of cfus, this seems to be an unlikely explanation for this case. On the other hand, the splicing effect could explain both the lower overall efficiency and the differences between the introns targeting *sacB*. Importantly, this does not exclude other effects from contributing to the observations and therefore further experiments would be required to identify the actual underlying cause. Especially sequencing of the genomic *sacB* locus for whole cultures prior to the sucrose selection could elucidate what fraction of the DNA actually carries the intron insert. Nonetheless, the efficiency of *sacB* disruption at position 491s was assumed to provide sufficient sensitivity for the selection of disrupted gene loci following the delivery of RNA with the MP.

3.3 Delivery of RNA into *E. coli* via MS2 maturation protein

3.3.1 Purification of N-terminally tagged MP

The next step in building a MS2-based delivery system for RNA, was to reconstitute the formation of the so called minimal infectious units *in vitro*. This step requires purified and soluble MP, however, the maturation proteins of the related bacteriophages Qβ and R17 are known to be highly insoluble when isolated [226,227]. Therefore, it was likely that this also applies to the MP of MS2, presumably caused by hydrophic residues required for the interaction with

the CP dimers [105,203]. In consequence, a variety of recombinant MS2 MP fusion constructs with diverse affinity and solubility enhancing tags were cloned. An overview of all the fusion constructs important for this thesis is given in Figure 12A. As a starting point, a MP variant with N-terminal hexahistidin-tag (His-MP) was used, for isolation of the MP through immobilized metal ion affinity chromatography (IMAC).

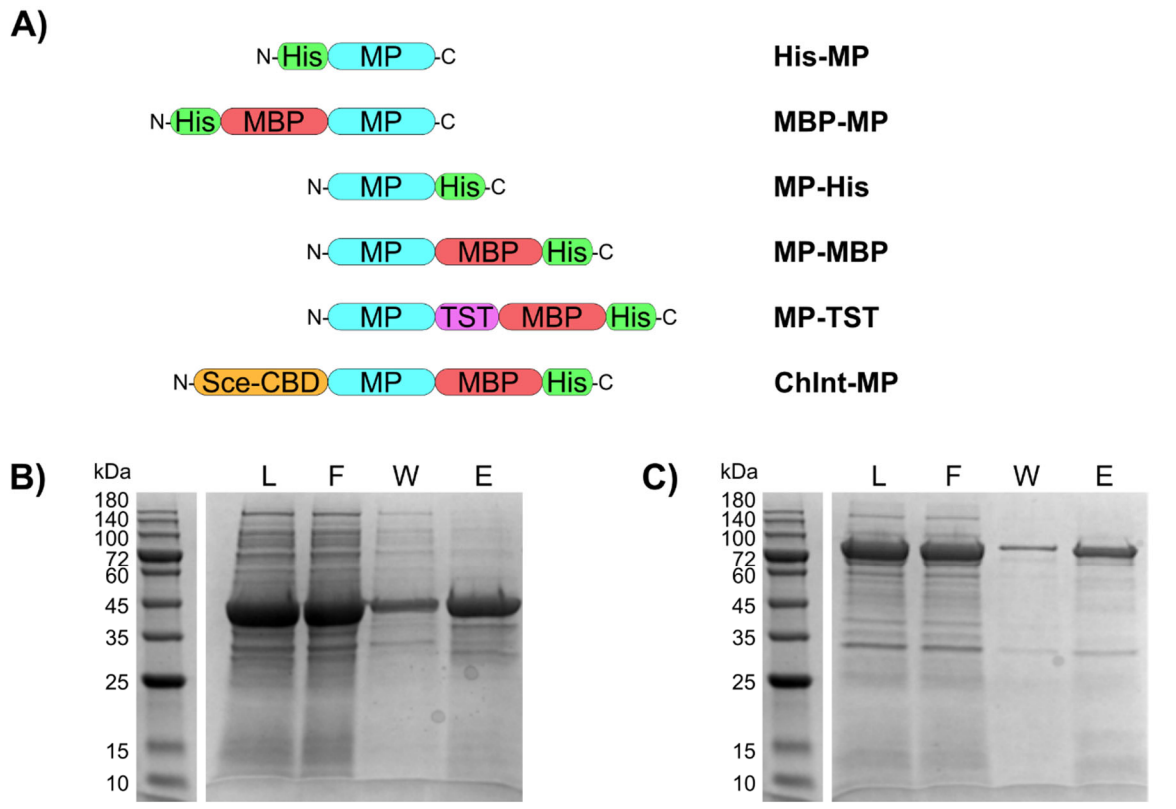


Figure 12 Recombinant MP fusion constructs and initial purification: A) Overview of the different fusion constructs of the MS2-MP (MP) used for this thesis, including the order of tags from the N- to the C-terminus, as well as how each variant is being referred to. His stands for the hexahistidin-tag for IMAC, MBP stands for the maltose-binding-protein-tag for solubility enhancement, TST stands for the Twin-Strep-tag[201] for a potential selection experiment and Sce-CBD stands for a chitin-binding-domain (CBD) embedded into a modified self-cleaving Sce VMA intein (Sce)[202]. **B)** and **C)** SDS-Page of samples taken during IMAC of soluble fractions of His-MP (**B)**) and MBP-MP (**C)**), respectively, after renaturation. The gel loadout is from left to right: loaded fraction (L), flowthrough (F), wash (W) and elution (E).

While being expressed at high levels however, this variant turned out to be severely insoluble, as it was exclusively found in the cell pellet after lysis (Supplementary Figure 3A). Therefore, the His-MP was extended with a maltose-binding-protein-tag (MBP-MP), a tag that is commonly used to enhance solubility of recombinantly expressed proteins [228–230]. But, as shown in Supplementary Figure 3A, this endeavour did not lead to an increase of solubility. Consequentially, the general purification strategy was changed from a direct IMAC-based

purification after cell lysis to the isolation of overexpressed MP from protein aggregates, also referred to as inclusion bodies (IB). This approach entails the initial purification of the IBs formed during overexpression, followed by denaturation to resolubilize the IBs, renaturation to allow refolding of the MP into its native conformation and finally an IMAC-based polishing step. Purification strategies without denaturing/renaturing steps are usually more favourable, as biological activity might be lost or aggregates reform under non-optimal refolding conditions[231,232]. However, the purification from IBs also has some advantages. For one, it does not require any optimisation of expression conditions that usually serve to increase solubility. Secondly, they can be easily isolated after cells lysis and consist mostly of the overexpressed protein[233–236]. The tested denaturing conditions are listed in Table 11.

Table 11: Denaturing Conditions for MP purification from inclusion bodies.

Condition	Urea [M]	SDS [%]	iPrOH [M]	Tris·HCl	pH
1	7	0	0	50 nM	7.5
2	7	1	0	50 nM	7.5
3	2	2	0	50 nM	7.5
4	2	0	6	50 nM	7.5
5	2	1	6	50 nM	7.5
6	7	0	0	50 nM	12.5
7	2	1	6	50 nM	12.5

Both for His-MP and MBP-MP, conditions 2 and 3, corresponding to a pH of 7.5, urea and SDS concentrations of 7 M and 1 %, or 2 M and 2 %, respectively, lead to a high proportion of protein in the soluble fraction after centrifugation (Supplementary Figure 3B, C). In case of MBP-MP, conditions 1, 6 and 7, corresponding to 7 M urea at pH 7.5, 7 M Urea at pH 12.5 and 2 M urea, 1 % SDS and 6 M iPrOH at pH 12.5, respectively, also proved to be suited for IB resolubilization. However, to keep the conditions between both preparations more comparable, only preparations denatured under condition 2 were used for the refolding. Condition 3 was omitted, as even though it did prove to work for denaturing purposes, the higher concentration of SDS caused issues at lower temperaures, due to precipitation of SDS.

To refold the denatured MP-preparations, a buffer was chosen for an initial refolding experiment, containing 0.5 M urea to prevent re-aggregation. In addition, the buffer was supplemented with 1 % polyacrylic acid, which was found to enhance solubility and prevent aggregation of the *Qubevirus* (Qβ) maturation protein[227]. This buffer composition already yielded highly enriched and soluble His-MP and MBP-MP (Figure 12B, C).

Following the refolding via dialysis and centrifugation, the solubilized MP-preparations were

polished using a single IMAC step, leading to highly pure His-MP and MBP-MP, respectively (Figure 12B, C). However, a considerable part of the recombinant MPs did not bind to the column, as can be seen by their intense band in the flowthrough fraction. This quite likely might be due to overloading of the Ni-NTA resin. But it might potentially also be a result of impeded binding of the affinity tag, either due to urea interfering with the binding itself or to low accessibility of the tag after partial misfolding.

3.3.2 Evaluating the formation of minimal infectious units with MP variants

The next step in reconstituting the MS2 infection module was to test the purified MP for biological activity. This was attempted by complexing genomic MS2 RNA (Roche) with the MP-fusion proteins and subsequently infecting *E. coli* susceptible to MS2 infection (*E. coli* F^+5695) and performing plaque assays with the infected cells. However, this did not lead to any formation of visible plaque.

To identify if this was a consequence of misfolded MP or some other cause, the plaque assay was repeated with His-MP and MBP-MP now being expressed from corresponding plasmids in the PURExpress (NEB) *in vitro* translation system and then supplemented with MS2 RNA (Figure 13A). As infectious units of MS2 can be made from genomic RNA *in vitro*[184] (Figure 13B, C), it was assumed that *in vitro* expression of active MP and formation of infectious units by subsequent addition of MS2 RNA should also be possible. Interestingly, even in this case, no plaque formation was observed. This indicated that there might be a different cause for the lack of activity, besides misfolding.

A likely cause for tagged proteins to lose partial or full activity can always be the tag itself, potentially interfering with proper folding, or by occluding important binding sites. Thus, plasmids encoding for the variants MP-His, MP-MBP and MP-TST were cloned, all harbouring the tags at the C-terminus. This allowed for testing if the position of the tag interferes with the activity. In addition to these constructs, the untagged wildtype MP, as well as His-MP as a reference and the MS2 coat protein as a negative control were tested for their ability to form infectious units (Figure 13D). As expected, wildtype MP did indeed lead to the formation of infectious units, as monitored by plaque assay. His-MP on the other hand did again not lead to any visible plaques, hinting that a N-terminal tag interferes in some way with infectivity. Most remarkably, all C-terminally tagged variants did exhibit biological activity to some extent. Furthermore, it seems that both MBP-tagged variants increased infectivity twofold (MP-MBP) or fourfold (MP-TST), as compared to MP-His.

The increase does not yet account for the difference in template input and resource requirements due to different sizes of the encoding plasmids (~18 % bigger for MP-MBP and MP-TST) and the expressed proteins (~100 % bigger in comparison to MP-His), respectively. Therefore, the effect might be higher under normalised conditions.

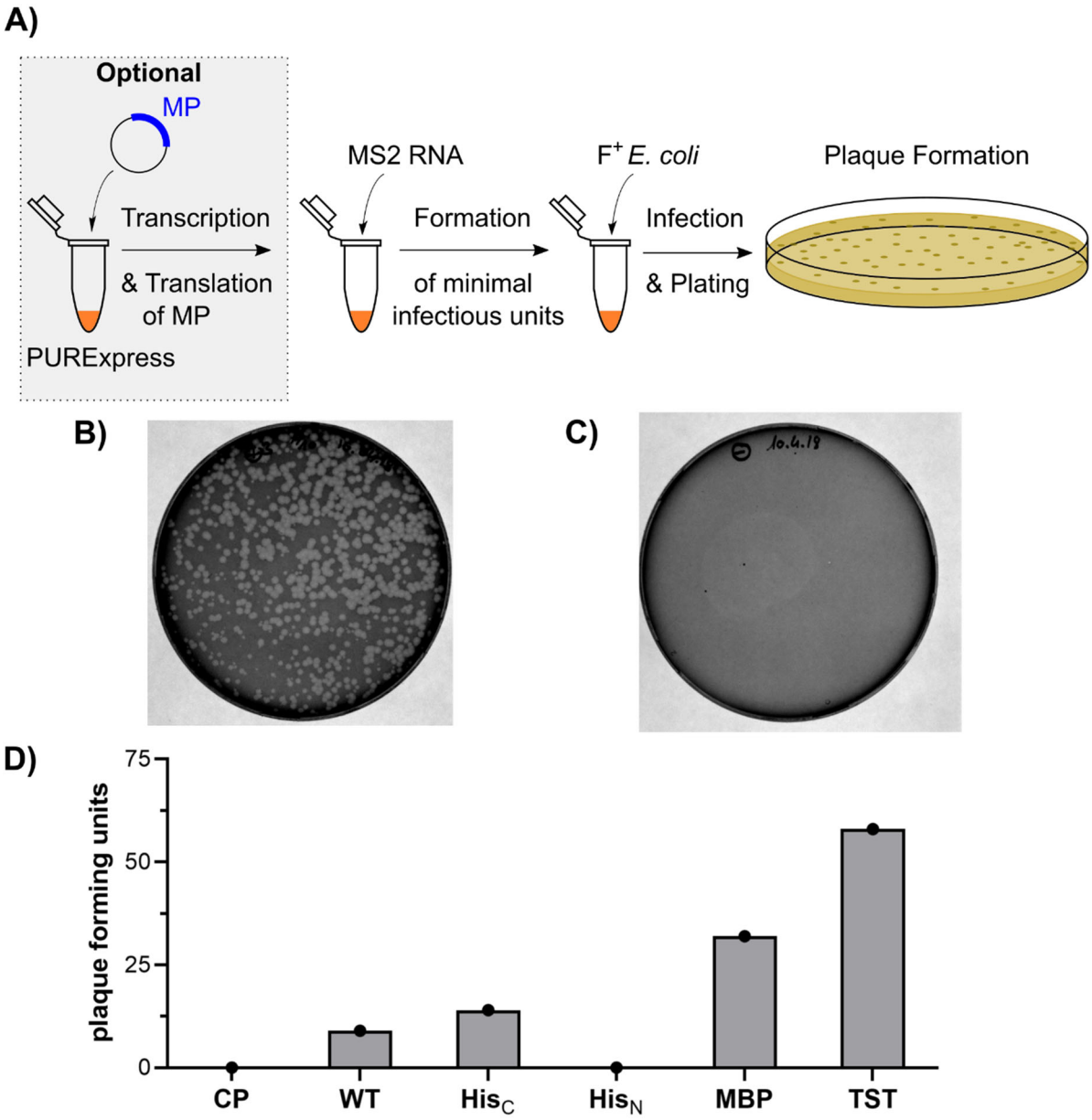


Figure 13 *In vitro* reconstitution of the MS2 minimal infectious unit: A) Scheme for the formation of minimal infectious units of MS2 *in vitro*. In an optional preparation step, the *mp* gene gets expressed in the PURExpress (NEB) *in vitro* translation system. Following this step, MS2 RNA is provided to allow the formation of infectious RNA-MP complex. If the optional step is omitted, incubation of MS2 RNA in PURExpress (NEB) will allow the expression of the native mp-gene, ultimately leading to the formation of the infectious RNP. Incubation of this RNP with MS2-susceptible *E. coli* F$^+$5695 leads to infection of these cells with MS2, which can be visualized by a plaque assay. In this assay, infected cells will lead to translucent spots, the so-called plaque, in an otherwise opaque, dense bacterial lawn. **B)** and **C)** Comparison of the bacterial lawn of cells infected with MS2 (**B)**) and cells that were not infected (**C)**). The bright spots in **B)** are translucent plaques caused by cell lysis and subsequent lack of bacterial (re-)growth. **D)** A comparison of the infectivity of different MP constructs, that were expressed in PURExpress (NEB) and subsequently complexed with MS2 RNA. The infectivity was measured as plaque forming units (pfu) per 12.5 µL reaction, supplemented with 150 ng of the corresponding plasmids. Following incubation at 37°C for 2.5 hours, 500 ng of MS2 RNA (Roche) were added, and the mixes incubated at room temperature for 15 minutes to allow for RNP formation. The tested proteins were MS2 coat protein (CP) as a negative control, untagged wildtype MP (WT), MP-His (His_C), His-MP (His_N), MP-MBP (MBP) and MP-TST (TST), following the nomenclature shown in Figure 12A.

However, as this experiment was only performed as a single proof-of-concept experiment, absolute quantification is not possible with the data present, and the observed differences might actually not be as substantial. Nonetheless, a higher solubility of the MPs tagged with MBP presumably increases the observed number of plaques and therefore is the most plausible cause of any observed effect.

3.3.3 The role of the N-terminus during infection

The next step was now to purify the C-terminally tagged MP variants and test if they can be isolated in an active form. However, in contrast to the high expression levels of N-terminally tagged variants, no expression was observed for MP-His, MP-MBP or MP-TST. It is known that specific residues at the N-terminus can target a protein for degradation. This is being referred to as the N-end rule and is a principle highly conserved in both prokaryotes and eukaryotes [237–240]. Within the N-terminus of the MP (Figure 14A, B), three amino acids can be identified that are presumably either directly destabilizing (F4 and L7) or can become destabilizing after modification (R2) [237]. While not directly at the N-terminus, proteolytic cleavage of the preceding amino acids might reveal either one of the three destabilizing residues and mark the MP for degradation [237].

Since the overexpression of His-MP was possible, this was assumed to be a highly likely cause for the lack of expression of the C-terminally tagged MP variants. Therefore, mutants of the C-terminally MP variants were cloned, where all three destabilizing residues were replaced with either a non-destabilizing histidine or a non-destabilizing alanine (MP(mut), Figure 14A). Indeed, these changes allowed overexpression of the corresponding MP variants (Supplementary Figure 4). However, as subsequent plaque assays revealed, they also completely abolished infectivity (Figure 14C). This does not come surprising, having a more closely look at how the N-terminus of the MP and the 3'-end of MS2 RNA interact in the mature MS2 virion (Figure 14B). As discovered by *Dai et al.* [105], the position of the N-terminus is important for the correct binding of the RNA, with a special significance of F4. This residue does directly interact with U3553 at the 3'-end of MS2 RNA, via RNA base stacking with its aromatic side chain [105]. To explore if other mutations might allow for overexpression, and maintain the formation of infectious units, several new mutants of MP-MBP were cloned (Figure 14A). These mutations were R2Q (MP(1a)), R2H (MP(1b), R2Q with F4I (MP(1a2)), R2H with F4I (MP(1b2)), R2Q with L7A (MP(1a3)) and R2H with L7A (MP(1b3)).

All MP variants carried a mutation of R2, as it was assumed that the most N-terminal destabilizing residue might also have the biggest impact on protein degradation. Additionally, the two mutants MP(1a2) and MP(1b2) served as tests for the effect of impeded RNA binding on infectivity.

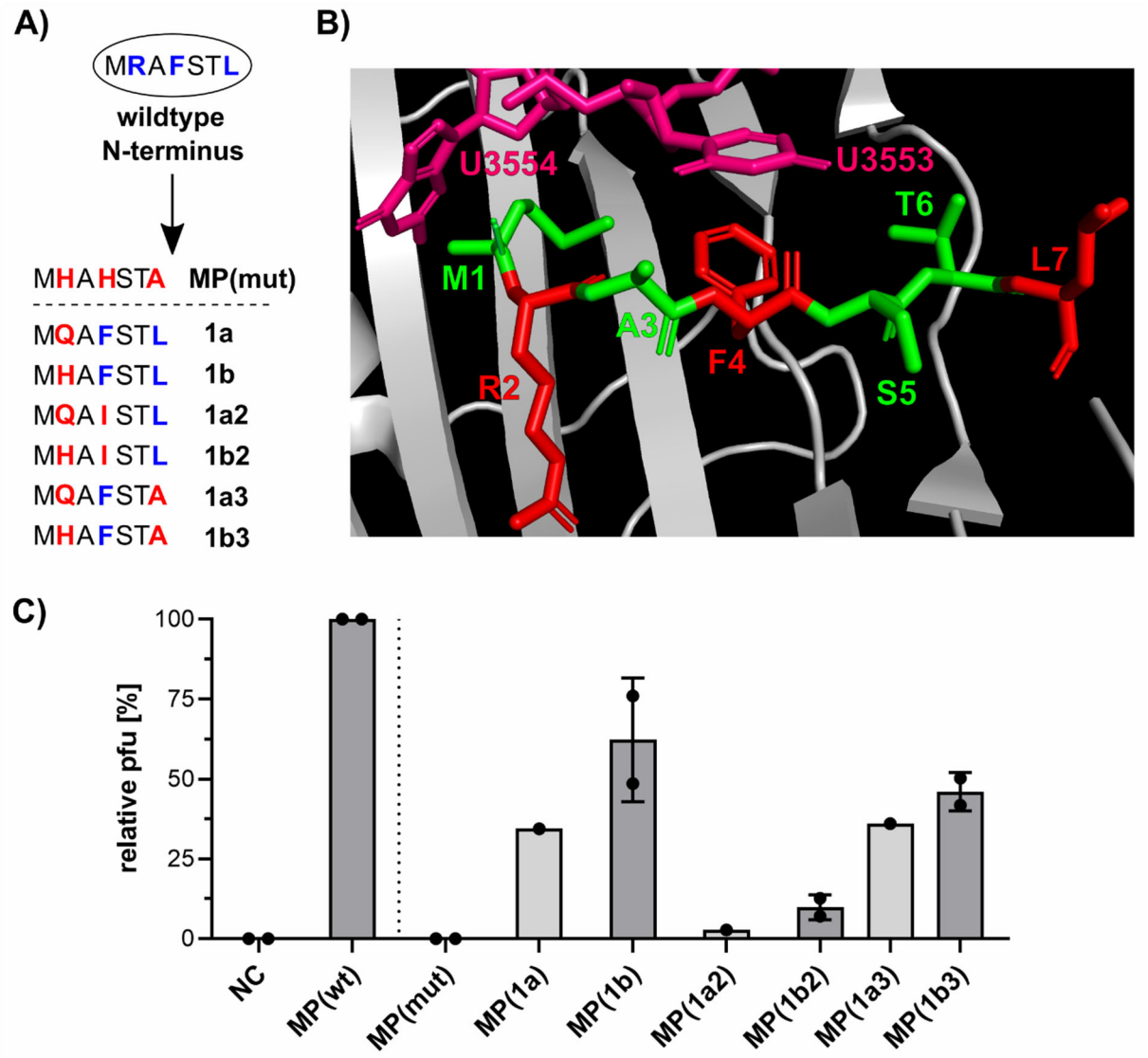


Figure 14 Analysis of N-terminal mutants of the MP: A) Overview of the first seven N-terminal amino acids of the MP and the mutant variants, in single letter abbreviation. Bold, coloured letters indicate the destabilizing amino acids (blue), or respectively, their mutated counterpart (red). Indicators next to the amino acid sequences show how each mutant is being referred to. **B)** Detail view of the N-terminus of the MS2 maturation protein and the closest nucleotides of the 3'-UTR of the genomic MS2 RNA, based on the cryogenic electron microscopy structure of MS2 RNA packaged into the viral capsid (PDB: 5TC1)[105]. The first seven amino acids of the N-terminus of the MP (grey) are depicted in sticks style with non-destabilising amino acids in green, destabilizing amino acids in red. Additionally, the two nucleotides closest to the N-terminus (U3553 / U3554, pink) are depicted as sticks as well. The image was prepared using open source PyMOL version 2.6.0a0[208]. **C)** Comparison of the relative infectivity of N-terminal mutants shown in panel **A)**. The MP variants were generated by programming 12.5 µL PURExpress (NEB) with 150 ng of the plasmids encoding the corresponding MP-MBP variant, followed by incubation at 37°C for 3 hours. Subsequent addition of 500 ng of MS2 RNA (Roche) and incubation at room temperature for fifteen minutes allowed formation of infectious units for plaque assay. The negative control (NC) was obtained by omitting the expression plasmid from the PURExpress (NEB) reaction mix. Infectivity, represented by pfu was normalised to the pfu counted for wildtype MP-MBP. All variants carrying the H2Q mutation (light grey: 1a, 1a2, 1a3) were only tested as single proof-of-concept replicate, all other variants (dark grey) were tested as biologically independent duplicates Error bars show the standard deviation of the mean, based on two independent, biological replicates.

In PURExpress-based plaque assays, all mutants harbouring the R2Q mutation (1a, 1a2, 1a3) seemed to produce less plaque than their corresponding counterpart with the R2H mutation (1b, 1b2, 1b3) (Figure 14C). This might indicate that a positive charge at position two, as is the case for arginine and histidine, might be more beneficial than a structurally similar but uncharged side chain, as is the case for asparagine.
However, for simplicity, MP(1a), MP(1a2) and MP(1a3) were not used for the second replicate. Regarding the overall effect of the R2 mutations, it seems that there is an observable negative effect on infectivity, compared to the wildtype MP, approximately reducing it by 25 – 60%. Based on the three plaque assays done for variants with the additional L7A mutation, there was no clear effect on infectivity compared to the mutants without this mutation. Presumably this is because the mutations of L7 are already too distant from the RNA binding site and do not alter the infectious conformation in a measurable way.
The most drastic effect on infectivity were observed for mutants that did not have a phenylalanine in position four of their N-terminus, MP(1a2), MP(1b2) and as expected MP(mut). While at least some negative effect was expected, the almost complete abolishment of plaque formation was surprising, considering that the interaction between F4 of the MP and U3553 of the RNA is only one of more than 20 distinct interactions between the MP and the MS2 RNA [105]. All in all, these results verify the importance of the correct amino acid at the N-terminus of the MP and puts special emphasis on the role of F4 on the infection potential of minimal infectious units of MS2 or potential synthetic RNAs.

3.3.4 Purification of active MP-MBP

Considering the importance of the N-terminus, the strategy for purification of active MP was changed a final time. This time, the MP was tagged C-terminally with an MBP-heptahistidine-tag to enhance solubility and allow for downstream polishing. N-terminally it was tagged with a self-cleaving *Sce* VMA intein, also harbouring an internal chitin binding domain (CBD) [202], protecting the MP from degradation during the overexpression, enabling chitin-based affinity purification and finally cleavage of the full N-terminal tag, generating the native N-terminus of the MP.
This MP variant (ChInt-MP) was successfully overexpressed in soluble form. Furthermore, the splicing of the *Sce*-intein-CBD-tag produced a protein of the expected size, in good purity (Figure 15A). This method has the advantage that due to the high affinity and binding strength of the CBD-tag, the wash conditions can be comparable stringent. Thus, most impurities can already be removed during a simple one step purification, as can be seen for the samples directly taken from the chitin resin (Figure 15A). However, especially at lower temperatures, the splicing of the intein from the desired protein can take some time. Even after 72 hours this reaction did not reach completion, but rather approximately 50 % cleavage (Figure 15A).

Possibly, this rate can be enhanced by a continuous supply of DTT, as the half-life time of this compound is rather limited under the purification conditions (approximately 12 hours) [241]. Interestingly, while both the non-cleaved protein and the cleaved intein-tag remain bound to the column after the elution step, a further unidentified protein co-purified in high amounts with the MP-MBP (Figure 15A). Considering the tendency of the MP to precipitate, it is likely that this co-purified protein is an *E. coli* chaperone, stuck to exposed hydrophobic residues of the MP [242]. Based on the estimated size, this chaperone might be GroEL (58 kDa), a chaperone known to co-purify with overexpressed proteins through hydrophobic interactions [242–244].

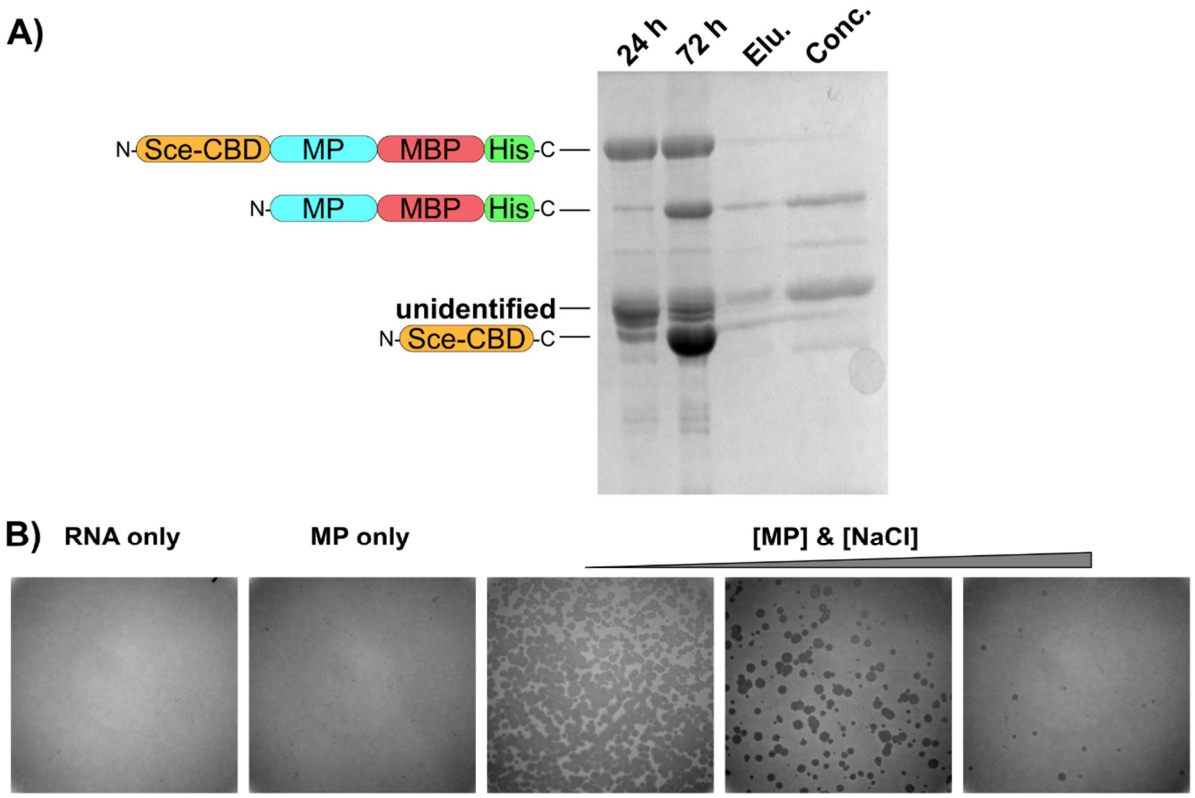


Figure 15 *In vitro* and *in vivo* characterisation of MP-MBP: A) SDS-Page of samples taken during the purification of MP-MBP. Samples of the chitin resin were taken 24 and 72 hours, respectively, after DTT was added to activate the intein splicing. The other samples were taken from the elution fraction (Elu.), as well as an up-concentrated fraction thereof (Conc.). The bands correspond to (from top to bottom) full length ChInt-MP (145 kDa), MP-MBP (87 kDa), an unidentified protein and the residual N-terminal inten-CBD-tag (58 kDa). **B)** Plaque assays with varying concentrations of MP-MBP (0.2 µM, 0.4 µM, 1 µM) and NaCl (50 mM, 100 mM, 250 mM). For the generation of infectious units, 1 / 2 / 5 µL of the concentrated MP-MBP preparation from **A)** were incubated in a binding buffer (20 mM HEPES pH 8.0, 10 mM $Mg(OAc)_2$) with 500 ng of MS2 RNA (Roche) for 30 minutes at room temperature. Subsequently these mixes were used to infect *E. coli* F^+5695 cells in standard plaque assays.

Nonetheless, this preparation proved to be biologically active, as formation of visible plaque occurred during plaque assays (Figure 15B). Neither MS2 RNA nor MP-MBP alone led to plaque formation. Only in combination, successful infection of *E. coli* was possible.

Interestingly, for this experiment the number of pfus inversely correlated with the input concentration of MP-MBP. In this case, the final reaction conditions were not adjusted to accommodate for the increasing concentration of the storage buffer of the MP (buffer DB1). Thus, the decrease in plaque count with increase in MP concentration can directly be correlated to an increase of NaCl concentration. This discovery matches well with the types of interaction between the MP and the MS2 RNA, proposed by *Dai et al.* [105]: Except for a single interaction, they are all based on charge-charge or dipole-dipole interaction and thus sensitive to disruption by increasing salt concentrations.

When the buffer conditions were kept constant, a clear dependency of infectivity on the concentration of MP-MBP was observed (Figure 16A). Albeit the clear variation between individual replicates, a general trend becomes apparent: While the observed pfu stay comparable for MP-concentrations 0.2 µM and 66 nM, they quickly drop once the concentration of the MP goes below 22 nM, and infectivity is almost completely lost at concentration below 0.8 nM. Looking at the molar ratios of RNA to MP, this trend translates to the following ratios: Below a ratio of 0.98 (RNA:MP), the plaque count stays stable, above this ratio, so at a concentration of the MP below 22 nM, the count quickly starts to drop. The binding of the RNA by the MP is essential for the infection. Thus, this suggests a dissociation constant potentially in the lower, two-digit nanomolar concentration range. However, as the MP was only isolated in low purity, protein contaminations, as well as RNAs were co-purified, the actual concentrations are unknown, and these results are more only speculative. In consequence, they require validation with a preparation of the MP-MPB of higher purity.

The substantial deviation between the three replicates was also observed in other experiments (data not shown). A possible explanation could be biological variations in the *E. coli* cultures, like the actual number of cells per assay, condition of the cells or abundance of displayed F-pili. But also technical variations, like the temperature and thickness of the LB-agar in which the infected cells were mixed into and which was then poured onto LB-agar plates, might have contributed. In combination, these factors could result in a drastic effect on the infection and growth of the *E. coli* cells. However, while the absolute numbers of pfu varied notably, general trends could still be observed throughout replicates.

Another way to evaluate the affinity of proteins to RNA are electrophoretic-mobility shift assays (EMSA). Therefore, an EMSA was performed with MS2 RNA, with the concentration of RNA kept constant, and the MP being titrated to the RNA in increasing concentration (approximately 0.8 nM to 0.2 µM) (Figure 16B). Unfortunately, no clear shift of the RNA was detectable. Considering that only one MP can bind per MS2 RNA strand, and that even the tagged MP-MBP only amasses to 7.6 % of the mass of the MS2 RNA, this result is not fully unexpected. However, even EMSAs with shorter RNA substrates did not show any protein-RNA interaction (Supplementary Figure 5). As already stated above, for a more reliable assertion regarding the

MP-RNA binding, these experiments must be repeated once a method was established that produces highly pure MP.

In general, while it was possible to reconstitute the minimal MS2 infection machinery *in vitro*, the overall efficacy, based on observed plaque formation and protein purity, did not satisfy the set expectations. The EcI5 intron RNA (491s) transcribed from plasmids could only successfully disrupt the *sacB* gene in ~10 % of the total population. In addition, an initial synthetic RNA genome might not be adapted to sufficiently replicate *in vivo*. Therefore, high transfection rates are crucial, to increase the chances to detect successful delivery of these RNAs.

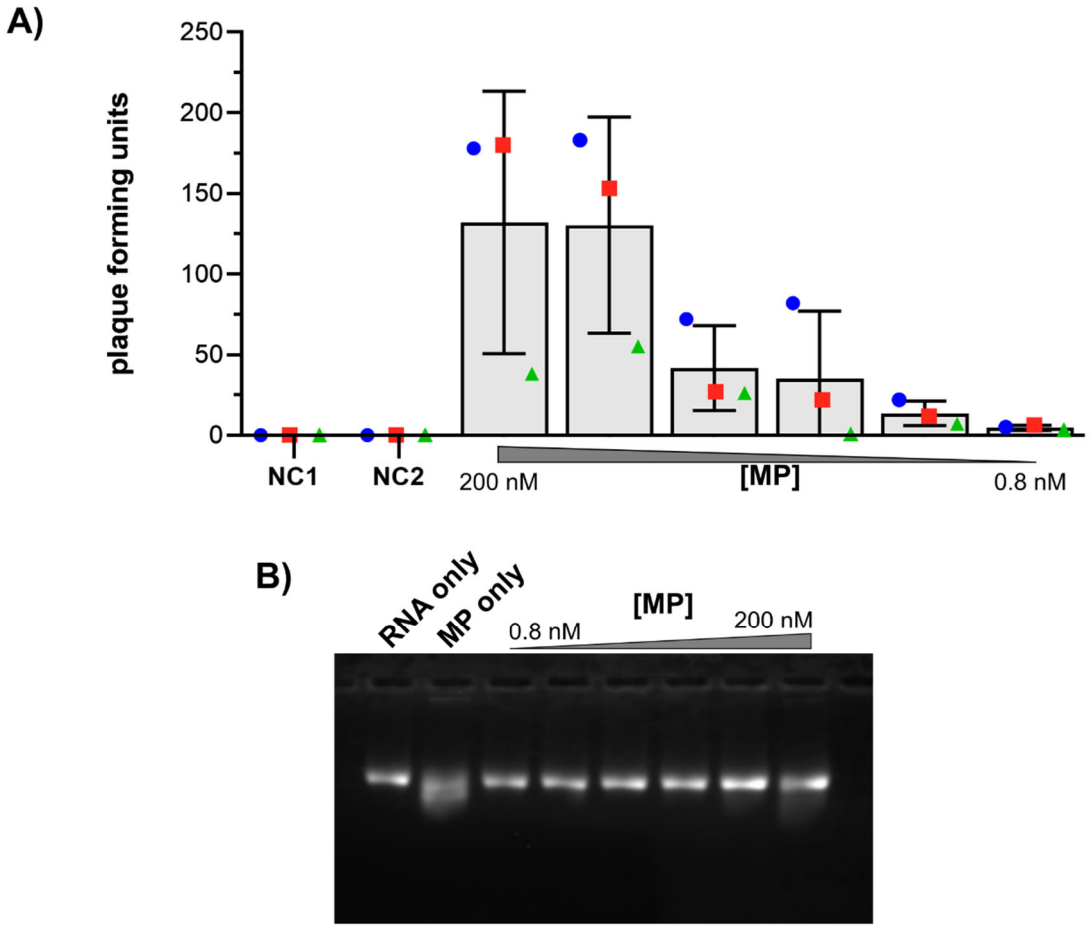


Figure 16 ***In vitro*** **reconstitution of minimal infectious units: A)** Comparison of the infectivity of minimal infectious units, depending on the concentration of supplied MP-MBP. For the plaque assays, 500 ng of MS2 RNA (Roche) were mixed with buffer compensated dilutions of MP-MBP (approximately 0.2 µM to 0.8 nM) in 1x PBB and incubated at room temperature for 30 minutes prior to infection. The three independent replicates are grouped by colour: replicate 1 = blue circle, replicate 2 = red square, replicate 3 = green triangle. As negative controls, cells were also incubated with either only MS2 RNA (NC1) or only MP-MBP (NC2). Error bars show the standard deviation of the mean, based on three independent, biological replicates. **B)** EMSA of genomic MS2 RNA with varying concentrations of isolated MP-MBP (approximately 0.8 nM to 0.2 µM). Samples were run on a 1 % TA agarose gel. Per sample, a 20 µL mixture was prepared with 25 nM MS2 RNA (H_2O for the MP only) and 2 µL of buffer compensated MP dilution in 1x PBB and 1x EMSA buffer.

However, the achieved infection levels are presumably too low, only ranging between 100 to 500 pfu per 500 ng of MS2 RNA, an RNA genome that is already adapted to *in vivo* conditions. For the planned selection experiment with the EcI5 intron or the delivery of an RNA genome, these levels probably need to be higher by at least two orders of magnitude. Furthermore, especially the co-purified RNA contamination make it seem unlikely to successfully select for RNA motifs that tightly and reliably bind to the MP. Thus, the idea of using this system for RNA delivery was put on halt in favour of alternative infection/transfection approaches.

3.4 Delivery of RNA via electroporation

A commonly used method for the delivery of DNA into cells is electroporation. This method has a broad spectrum of organisms that it can be applied to, ranging from a variety of bacteria [245,246], over yeasts [247,248] to mammalian cells [249,250]. Thereby, an external electric field presumably leads to a reversible and only temporary formation of pores within the cell membrane. This allows for the cellular uptake of low molecular weight molecules by diffusion, while high molecular weight molecules are internalized by processes based on electrophoreses [246,250]. In general, this method has the advantages of being highly efficient, based on high transformation/transfection rates, as well as low material requirements for making cells competent for electroporation [189,246]. Furthermore, this method also has been shown to be applicable for the delivery of RNA to yeast [251,252], mammalian cells [250,253] and, less commonly found in published literature, to bacteria like *E. coli* [254].

However, a downside to the commonly used protocols for the preparation of electrocompetent cells is the requirement for continuous cooling of all cultures, buffers, and consumables during the process. Deviation from this can already lead to noticeable reductions in transformation efficiencies, thus potentially causing relevant variation between individual cell preparations. *Tu et al.* describe a method for the preparation of electrocompetent *E. coli*, that circumvents this problem, by preparing the cells at room temperature [189]. Starting from this protocol, several variations of it, as well as commonly used protocols for the preparation at <4°C were benchmarked for their transformation efficiencies (Table 12, Figure 17A). This work was done by *Alexander Floroni*, in the context of his master thesis, under the supervision of the author of this thesis.

The standard room temperature protocol (RT1), as well as all variations, exhibited high transformation efficiencies at around 10^7 cfus per ng of transformed plasmid. In addition, the overall variability between individual replicates was lower compared to the protocols Cold2 and Cold3, and similar to Cold1. The biggest deviation for the RT protocols was observed for RT3 with less than one order of magnitude, while the deviations for Cold 3 and Cold2 were above one, or respectively, two orders of magnitude around their respective mean. From the protocols at <4°C, Cold1 and Cold3 performed best with efficiencies between 10^6 and 10^7 cfu/ng, while

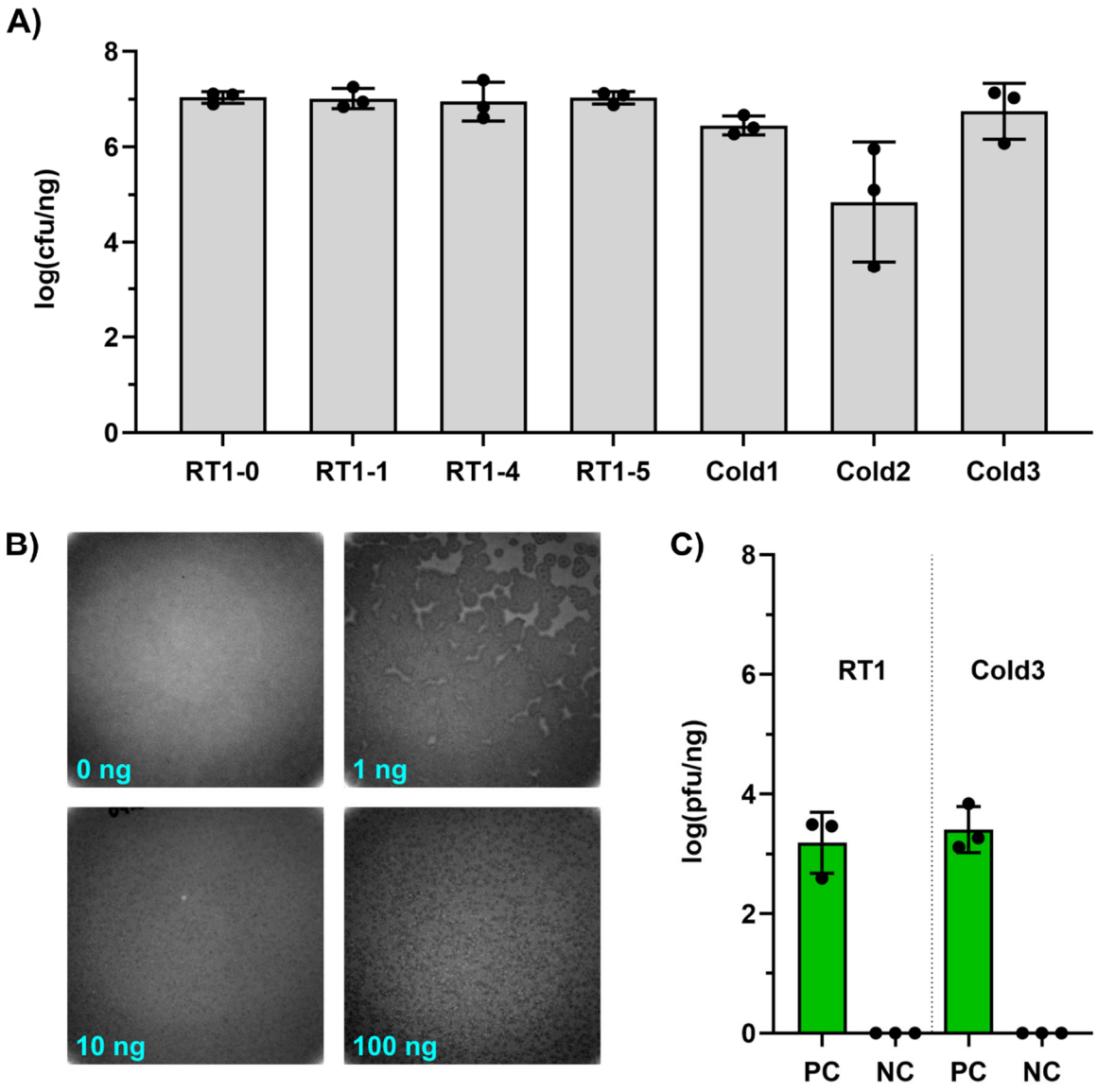


Figure 17 Electroporation of nucleic acids: A) Benchmarking of the protocols for preparing electrocompetent *E. coli*. The individual protocols are listed

Table **12**. RT1-5 refer to protocols for the preparation at room temperature, Cold1-3 to protocols for the preparation at <4°C. For the assay, cells were made electrocompetent, transformed with 1 ng pMini-Tnp(empty) (2.3 kb) and plated in dilutions on LB-agar plates containing Carbenicillin. Transformation efficiencies were calculated as the log of the number of cfus per ng of transformed DNA. Error bars depict standard deviation of the mean, based on three independent, biological replicates. **B)** Images taken from plaque assays of *E. coli* transfected with MS2 RNA (IVT). Cells were made electrocompetent at room temperature, electroporated with 0 / 1 / 10 / 100 ng of *in vitro* transcribed MS2 RNA. Subsequently, infected cells were used for a standard plaque assay. **C)** Comparison of transfection efficiencies between the protocols RT1 and Cold3, based on plaque assays of cells that were electroporated with 0.01 ng of MS2 RNA (PC). As a control, electrocompetent cells were mixed with the same amount of RNA, but not electroporated (NC). Transfection efficiencies were calculated as the log of the number of pfu per ng of transformed DNA. Error bars depict standard deviation of the mean, based on three independent, biological replicates. Data for **A)** and **C)** was taken from the master thesis of *Alexander Floroni*.

Cold2 only produced 10^5 cfu/ng on average. Considering the range of error, Cold1 and Cold3 are comparably efficient as the RT protocols, albeit less robust.

The next step was to test these methods for the delivery of MS2 RNA into *E. coli* F^+5695. As shown in Figure 17B, efficient transfection with MS2 RNA was possible. Furthermore, the number of pfus scales with the amount of transfected RNA. There were still individual plaques visible for 1 ng of RNA. However, with increasing amounts of transfected RNA, the opaque bacterial lawn with translucent spots turned into a translucent lawn of lysed cells with small, single colonies. While plaques at lower concentrations were comparably big, the size of visible plaques inversely scaled with the RNA input. Together, these findings suggest that at high multiplicities of infection, lysis of infected cells occur at a rate too fast compared to cell growth. Thus, the diffusion of MS2 virions is disabled, as a too fast lysis prevents the formation of a continuous bacterial lawn required for the spreading. This is limiting the observed size of plaques. Additionally, this enables some cells to potentially escape infection by MS2, leading to the growth of small, single colonies.

Having shown that the RT1 protocol works for the transfection of MS2 RNA, this protocol was compared to Cold3. Cold3 was chosen as it performed similarly to the RT protocols, but also was different enough to not be considered a simple variant of RT1. For both protocols, the addition of MS2 RNA to electrocompetent cells alone did not lead to formation of visible plaque (Figure 17C). The formation of plaque only occurred when cells were mixed with the RNA and subsequently electroporated.

Table 12: Protocols for the preparation of electrocompetent *E. coli* cells.

Protocol	OD_{600}	Volume Culture [mL]	Wash Steps	Glycerol [%]	Final volume [µL]
RT1	**0.2**	**1.4**	**2**	**0**	**30**
RT2	**0.2**	**10**	**2**	**0**	**30**
RT3	**0.4**	**1.4**	**2**	**0**	**30**
RT4	**0.6**	**1.4**	**2**	**0**	**30**
Cold1	**0.4**	**1**	**3**	**0**	**40**
Cold2	**0.4**	**25**	**3**	**10**	**50**
Cold3	**0.6 – 0.7**	**20**	**4**	**8.7**	**50**

Based on data gained from this experiment, there were no clear differences between RT1 and Cold3. In comparison to the electroporation of DNA (10^7 cfu/ng), however, absolute numbers

were lower by approximately four orders of magnitude (10^3 pfu/ng).
While this might appear as a drastic reduction of the overall efficiency at first, there are several reasons as for why the transformation of DNA and RNA cannot directly be compared. For one, plasmidic DNA is less prone to degradation by nucleases than linear, single stranded RNA. Secondly, the process of infection and lysis by MS2 is potentially more complex than the maintenance of a plasmid that provides a selection advantage. In theory a single copy of both should be enough to get a cfu/pfu. But the difference of *in vivo* stabilities might require a cell to be infected with more than just one copy of MS2, to actually lead to formation of plaque. Therefore, the real transformation efficiencies for RNA and DNA might be more similar. In summary, a method was identified for the robust and efficient delivery of RNA, independent of MP/F-pili system. Furthermore, the presented results are well in accordance with previous findings for the related bacteriophage Qβ [254].

3.5 Reprogramming MS2 through random insertion mutagenesis

Designing a MS2 derived replisome based on the MP/RNA delivery system and a MS2 replicase based replication module can be considered a constructive approach. In contrast, the approach described in the following would be reductive, using the native MS2 genome architecture and reducing its native functions.
In theory, an easy route to a virus derived RNA replisome would be to knock-out all toxic genes or insert/replace them with any gene of interest. However, many RNAs encode information not only in their nucleotide sequence, but also through secondary structures [128,255], tertiary structures [256–258] and long-range interactions [259–261]. The translational control of MS2 genes, as well as the control of RNA replication, are known to be extensively coordinated through RNA structures [127,128,137,180,262]. Consequently, insertions into the MS2 genome or modifications thereof might have detrimental effects, as these might easily disrupt structural elements, causing a loss of information or function.
Rational design of corresponding DNAs, harbouring the desired insertion and serving as IVT templates, can circumvent or reduce these issues. However, this approach is limited by the available information and therefore prone to information biases, as well as potentially low feasibility of high-throughput screening of modified RNAs. Alternatively, random insertion mutagenesis can produce large libraries of mutants with the desired gene cassette randomly inserted at any position within the target DNA. This approach enables the screening of a drastically increased number of positions, potentially compensating the information bias of a purely rational design strategy.
The transposition system of bacteriophage Mu was chosen for the random insertion mutagenesis. This system consists of the recombinantly purified MuA transposase, as well as a gene cassette encoding for a selection marker (Figure 18A).

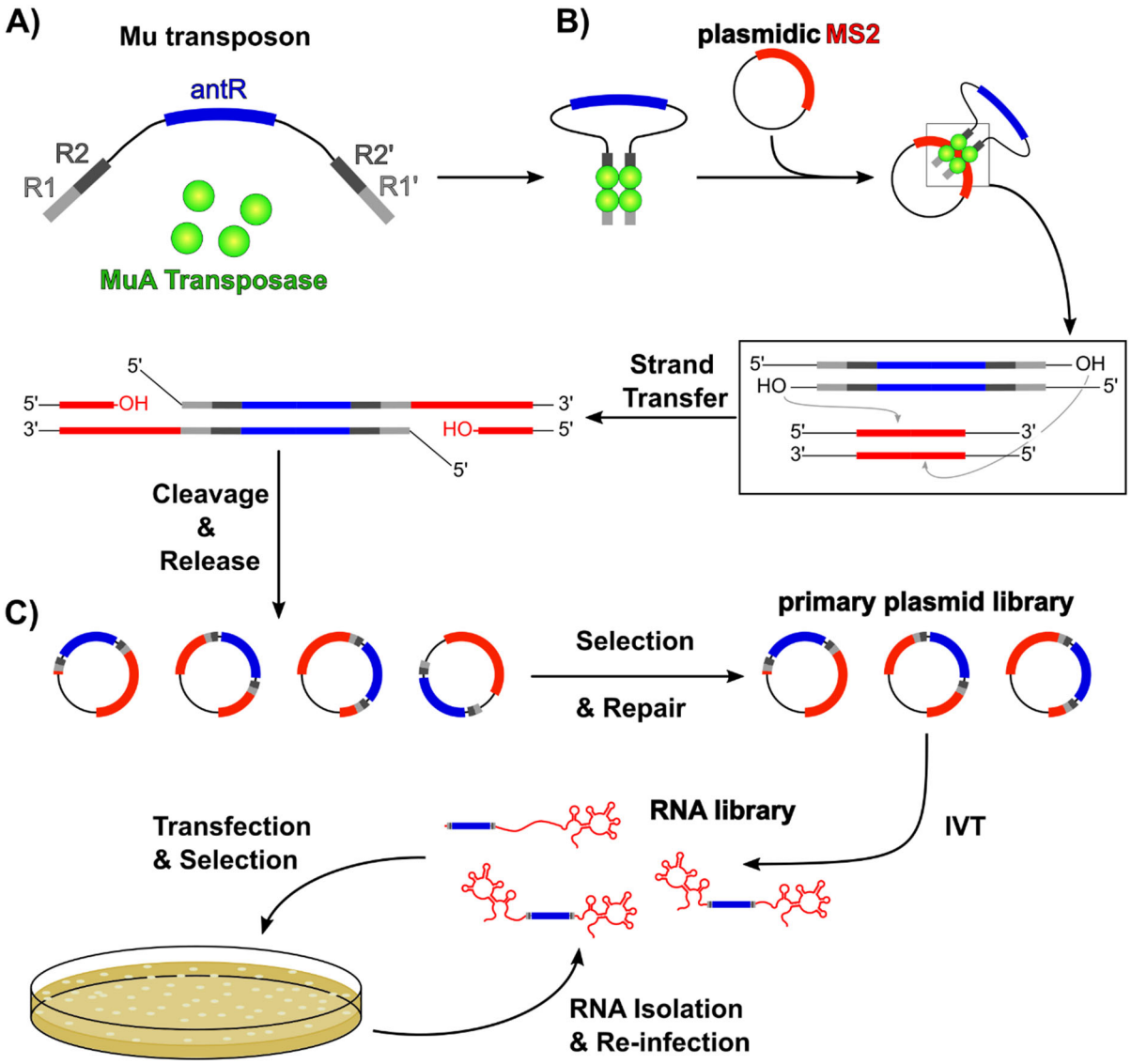


Figure 18 Mechanism of random insertion mutagenesis and library preparation using MuA transposase: A) The Mu-phage derived system for random insertion of a cassette into a target DNA consists of the MuA transposase, as well as a synthetic Mu-transposon. The transposon encodes for an antibiotic resistance gene (antR) flanked by two, bipartite recognition sites for the MuA transposon. These recognition sites were built by combining MuA binding sites R1 and R2, or respectively their complementary counterparts R2' and R1', each with a short 5'-overhang and a free 3'-OH. **B)** The binding of four MuA monomers to the transposon leads to the formation of the active transposome. Upon binding of the target DNA, in this case a plasmid encoding for the MS2 genome variants, the strand transfer reaction takes place. Thereby, the two free 3'-OH of the transposon randomly insert into the two strands of the plasmid within 5 bp distance. Subsequent cleavage of the residual 5'-overhangs of the transposon by MuA, produces the final mutant with 5 bp single stranded gaps at each end. **C)** Transformation of the initial pool of insertion mutants enables repair of the ssDNA gaps by *E. coli* DNA repair pathways, and in parallel can be used to select for mutants that have both selection markers (transposon & plasmid). These clones make up the primary plasmid library. Using this library as a template for *in vitro* transcription generates the final RNA library. Transformation of these RNAs into cells, with subsequent selection for the antibiotic resistance (transposon) allows the identification of self-replicating RNA scaffolds that encode for the selection marker. In an iterative process, based on RNA isolation and re-infection, these scaffolds can potentially be optimised by *in vivo* evolution.

The cassette used for this approach is based on the commercially available Mutation Generation System Kit (ThermoFisher Scientific). This cassette consists of a promoter for the expression in prokaryotes of the resistance genes against Kanamycin (*kanR*), or respectively Chlor-amphenicol (*camR*), embedded between two bipartite recognition sites for MuA transposase (R1R2, R2'R1') with short 5'-overhangs [205,263,264]. In a first step, four MuA monomers bind to the recognition sites on the gene cassette, thereby forming the active transposome (Figure 18B). Subsequently, the binding of the transposome to a plasmid encoding for the MS2 genome, allows for the strand transfer event to occur. Thereby, the MuA tetramer uses the free 3'-OH of the gene cassette to mediate a transesterification resulting in the cassette incorporated into the target plasmid. Finally, cleavage of the residual 5'-overhangs by the MuA tetramer and release (during DNA clean-up) of the DNA produces an initial pool of nicked insertion mutants [205,264,265].

Transformation of this pool into *E. coli* enables *in vivo* repair of the nicked plasmids, as well as selection for mutants harbouring the insert outside of the plasmidic selection marker (Figure 18C). The thus created primary DNA library can then be used as a template for IVT of the actual RNA library. Following the transfection of cells with these RNAs, selection of the selection marker of the transposon can be used to identify insertion mutants, capable of both *in vivo* replication and expression of a gene of interest. Furthermore, this approach can be expanded to optimise functional mutants through iterative selection of clones by adding RNA isolation and re-transfection to it.

For the reprogramming of MS2, three variants of MS2 were subjected to random insertion mutagenesis using transposons either encoding for *kanR* or *camR*, with varying success (Table 13).

Table 13: Results for the generation of random insertion libraries for three MS2 variants.

Variant / Antibiotic Resistance	Clones • 10^3	Max x-fold Coverage
MS2(wt) / Kanamycin	**7.2**	**1.01**
MS2(wt) / Chloramphenicol	**26**	**3.64**
MS2(qdef) / Kanamycin	**1.0**	**0.14**
MS2(qdef) / Chloramphenicol	**2.3**	**0.32**
MS2(ddef) / Kanamycin	**2.8**	**0.39**
MS2(ddef) / Chloramphenicol	**4.8**	**0.67**

The three variants were wildtype MS2 (MS2(wt)), a variant harbouring a defective CP mutant [203] and disrupted start codon of the *lysis* gene (MS2(ddef)), as well as a variant with additional

disrupted start codons for the *maturation* and *replicase* genes (MS2(qdef)). MS2(ddef) served as the actual scaffold for the generation of a self-replicating RNA that provides antibiotic resistance. MS2(qdef) served as a replication defective control to evaluate background resistance to the used antibiotics. MS2(wt) was used for an alternative approach. Here, instead of using antibiotic resistance for selection, self-replicating mutants that maintained their infectivity and lethality after random insertion mutagenesis were supposed to be screened for with plaque assays. Since in this case no antibiotic concentrations would have been to be optimised, this was regarded as a viable way to identify functional insertion mutants.

The theoretical minimum number of clones required to cover the full genome of MS2 (3569 nt) is 7136. This is because the insertion is random and can not only occur in the sense (genomic) orientation, but also antisense (antigenomic) orientation of the MS2 coding sequence. In case of MS2(wt), a round of insertion mutagenesis of *kanR* (1.01) and *camR* (3.64), respectively, both produced enough clones to theoretically cover the full genome (Table 13). For MS2(qdef), the potential coverage was substantially lower (0.14, 0.32), but considering its application as a control sample, this was sufficient. Finally, for MS2(ddef) the number of clones was lower than required for full coverage (0.39, 0.67), but still potentially high enough to include promising mutants.

To ascertain, if the generated libraries consisted of highly diverse clones, the primary plasmid libraries were digested with restriction enzymes ApaLI and SspI. The combination of the restriction enzymes produced specific cleavage patterns depending on the heterogeneity of the libraries. Indeed, as can be seen for the digest (Supplementary Figure 6), all libraries were majorly made up by clones with the insert inside the MS2 region. Additionally, all libraries were also found to be highly diverse, as a digest with the combined restriction enzymes produced a smear, corresponding to cleavage fragments of the MS2 region with a broad size distribution. However, while only seemingly visible for the three variants with the *camR* insert, all PCRs of the corresponding IVT templates also produced significant amounts of the insert-free MS2 variant. Thus, the alternative approach using MS2(wt) was not continued, as it was impossible to fully separate mutant MS2 from infectious wildtype MS2.

Nonetheless, RNAs of MS2(ddef) and MS2(qdef) with *kanR* and *camR* inserts, respectively, were electroporated into *E. coli* cells. Subsequently, the cells were plated on plates with varying antibiotic concentrations, to find optimal screening conditions. For all tested conditions, no difference in cfu was observed between the actual library and the control. This suggests that the tested conditions were not allowing for the expression of the antibiotic resistance or the replication and persistence of the transfected RNA. A potential absence of replication was assumed to be especially likely for two reasons. First, the theoretical and real coverage of positions was below one-fold. The *replicase* gene and the UTRs, which are essential for replication, make up 54.3 % of the total genome. Thus, it is highly possible that a relevant

portion of the mutants were replication deficient or attenuated due to insertion into the UTRs or *replicase* gene. Second, the MuA recognition sites are also inserted and present in both sense and antisense orientation. Therefore, they potentially form a 50 nt long double stranded RNA stem. As has been shown for the Qβ phage, double stranded regions can diminish replication efficiency or, if long enough, abolish it completely [266]. While this approach appeared to be quite promising, the test for *in vitro* replication of the insertion mutants prior to another *in vivo* evaluation was considered necessary.

3.6 *In vitro* characterisation of the MS2 replicase subunit

3.6.1 Reconstituting the replicase holo complex

The first step to testing the random insertion mutants for *in vitro* replication was to purify the MS2 replicase subunit (MS2rep). Therefore, the *rep* gene was cloned into pBAD33 expression vectors and purified through IMAC using either a N-terminal or a C-terminal His_6-tag. Furthermore, the replicase subunit of the related bacteriophage Qβ (Qβrep) was isolated as well, as potential back-up replication system which was already characterised in great detail [267]. Additionally, different buffer conditions (Mg^{2+} or EDTA) in the purification buffer were tested. In all cases, neither the position of the tag, nor the buffer composition had a visible effect on the protein preparation after a single IMAC step (Supplementary Figure 7). However, in a direct comparison between the two replicases, several differences can be observed.

For one, Qβrep was copurified with all known essential co-factors, ribosomal proteinS1 and elongation factors EF-Tu and EF-Ts [267,268]. MS2rep only co-purified with S1 and EF-Ts [100], but not with EF-Tu, which is an essential co-factor for MS2rep [66]. Second, the Qβrep complex was functional, as replication of RQ135 RNA, a known Qβrep substrate [269] was possible *in vitro* in trial experiments (Supplementary Figure 9, Figure 25A). In contrast, the MS2rep complex was not able to amplify $[F30\text{-}Bro(+/-)]_{UTRs(+/-)}$ RNA *in vitro* under comparable conditions [180].

However, as *Weise et al.* were able to show that MS2rep is active in the commercial translational PURExpress (NEB) system, if translated *in situ* from a corresponding mRNA [180]. Therefore, this system was used as a starting point for the *in vitro* reconstitution of a functional MS2rep replication system. The $[F30\text{-}Bro(-)]_{UTRs(-)}$ RNA was chosen as an initial replication template. Synthesis of $[F30\text{-}Bro(+)]_{UTRs(+)}$ from this template enables detection of replicase activity through a fluorescence signal. This signal is the result of binding of the fluorogen DFHBI-1T (Figure 19A) to the broccoli aptamer. Indeed, activity of MS2rep was detected in the PURExpress system, when $[F30\text{-}Bro(-)]_{UTRs(-)}$ was present (Figure 19B). While this was promising data, the proprietary nature and therefore unknown exact composition of the PURExpress system made a further reduction impossible. Changing to the homebrew PURE 3.0 [188] as the *in vitro* translational system, proved to be a viable alternative. Albeit the observed activity was lower compared to the PURExpress system (Figure 19B), it still provided all

essential co-factors. Consequentially, the well-defined composition of it enabled a further reduction, to identify a minimal MS2rep replication system.

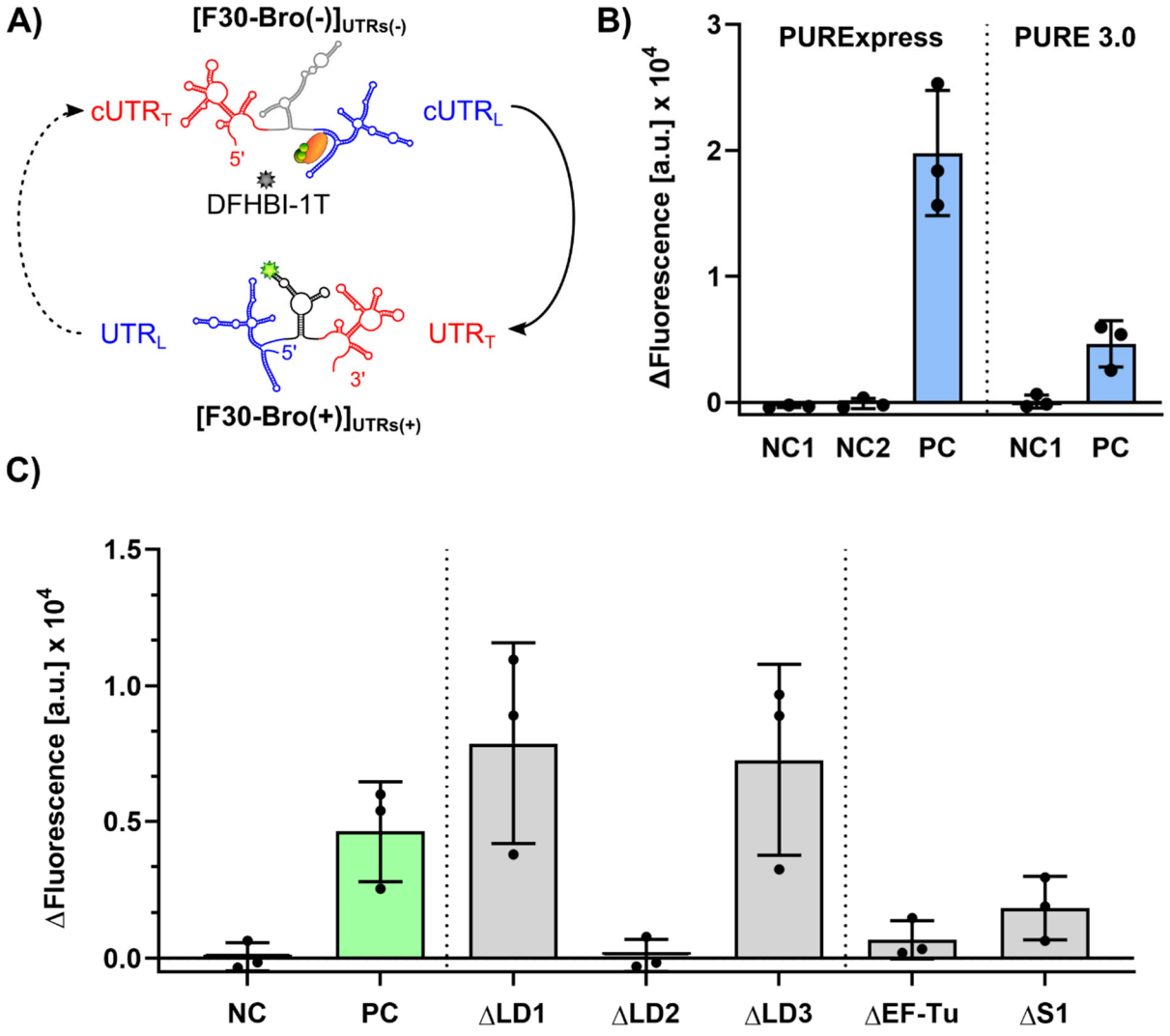


Figure 19 Synthesis of [F30-Bro(+)]$_{UTRs(+)}$ in PURE systems: A) Scheme of *in vitro* transcription of [F30-Bro(+)]$_{UTRs(+)}$ from [F30-Bro(-)]$_{UTRs(-)}$ by the MS2replicase holo-complex. The [F30-Bro(+)]$_{UTRs(+)}$ construct consists of the broccoli aptamer attached to the F30-stem [182], embedded between the 5'- (UTR_L, blue) and 3'-UTR (UTR_T, red) of the genomic MS2 RNA. The [F30-Bro(-)]$_{UTRs(-)}$ is the corresponding (-)-strand, with the complementary UTRs ($cUTR_L$ blue, $cUTR_T$ red) [180]. The MS2replicase subunit (orange), bound to its co-factors (green circles), transcribes the input [F30-Bro(-)]$_{UTRs(-)}$ into the complementary [F30-Bro(+)]$_{UTRs(+)}$. After folding of the nascent strand, the non-fluorescent DFHBI-1T (grey star) intercalates into the broccoli aptamer, thereby becoming fluorescent (light green star) enabling the fluorescence readout. **B)** Fluorescence change of [F30-Bro(+)]$_{UTRs(+)}$ in commercial PURExpress and homebrew PURE 3.0 after incubation at 37°C for 30 minutes. Reactions contained either 150 nM [F30-Bro(-)]$_{UTRs(-)}$ (NC1), 150 nM RNA and 300 nM purified MS2rep (PC), or only 300 nM MS2rep (NC2). Error bars indicate standard deviation for three technical replicates. **C)** Fluorescence change for depletion assay for homebrew PURE 3.0 programmed with 50 nM [F30-Bro(-)]$_{UTRs(-)}$. Reactions were assembled as in **B)**, but without the indicated component. The two controls were mixed with all components (PC), or respectively, all except MS2rep (NC). Error bars indicate standard deviation for three technical replicates.

The PURE 3.0 system consists of five individual components. LD1, LD2 and LD3 contain 30 out of 31 translation factors of *E. coli*. The enzyme mix combines ribosomes, EF-Tu and enzymes for creatine phosphate based NTP regeneration. Finally, the energy mix provides tRNAs, NTPs, amino acids and further low molecular weight components. Considering the homology between MS2rep and Qβrep (30.5% sequence identity / 43.2% sequence similarity [100]), as well as previous findings about MS2rep [66], the PURE 3.0 components were simplified. In conclusion, the enzyme mix was replaced with individual preparations of EF-Tu and S1, the energy mix was reduced to NTPs, $MgCl_2$, and DTT.

The adapted PURE 3.0 system, PUREred, was than deployed for further *in vitro* assays. In addition to testing only for activity of the full PUREred mix, a permutation depletion of the individual components was also investigated. In accordance with the expectations, MS2rep was still active in this reduced PURE system (Figure 19C). Furthermore, two of its components were identified to be non-essential (LD1, LD3), while the other three components (LD2, EF-Tu, S1) were required for full replicase activity. Omittance of LD1 and LD3 seemingly increased activity. This was presumably due to excluded-volume effects that increased formation of replication-inactive double stranded RNA, and/or altered RNA folding. While activity was almost fully abolished in the absence of LD2, depletion of EF-Tu and S1, respectively, still allowed for low level activity. In case of S1, a potential cause for this might be the co-purified S1, that however might only be present in sub stochiometric amounts. For EF-Tu, a contamination or copurification with the LD1-3 fractions seems possible. All in all, this experiment identified LD2, EF-Tu and S1 as the essential fractions.

LD2 consists of eight different *E. coli* proteins, elongation factor EF-Ts, Methionyl-tRNA formyltransferase (MTF), the aminoacyl-tRNA-synthetases AlaRS, AsnRS, IleRS and PheRS (α and β subunit), as well as initiation factors IF1 and IF3 [188]. To further narrow down the number of essential co-factors, these eight proteins were individually purified. All were obtained in high puritiy (Supplementary Figure 8), although EF-Ts co-purified with substantial amounts of EF-Tu. This supports the assumption of why depletion of EF-Tu did not fully abolish replicase activity.

In analogy to the PUREred depletion experiment, the individual components were omitted from a reconstituted LD2 mix and the effect on replicase activity was analysed (Figure 20A). The full reconstituted LD2 mix enabled replicase activity, although at a lower level as expected from the experiment PUREred depletion experiment (Figure 19C). Omittance of MTF, AsnRS or PheRS did not substantially affect the replicase activity. In case of IF1, AlaRS and IleRS, a moderate decrease (IF1, AlaRS), or respectively, increase (IleRS) was observed. The strongest effects were observed for EF-Ts and IF3, where omittance lead to a complete loss of activity (EF-Ts), or respectively, to an approximately threefold increase of activity (IF3). Although EF-Ts co-purified with the replicase subunit, this was expected as it was already

found to be an essential co-factor and the co-purified amounts were likely to be insufficient. For IF3, this finding was unexpected, but it indicates that IF3 might be a potential inhibitor of MS2rep activity. Additionally, it could explain the reduced activity with the full reconstituted LD2 compared to the PURErred LD2, as the concentration of IF3 in this experiment was presumably higher and inhibition therefore stronger.

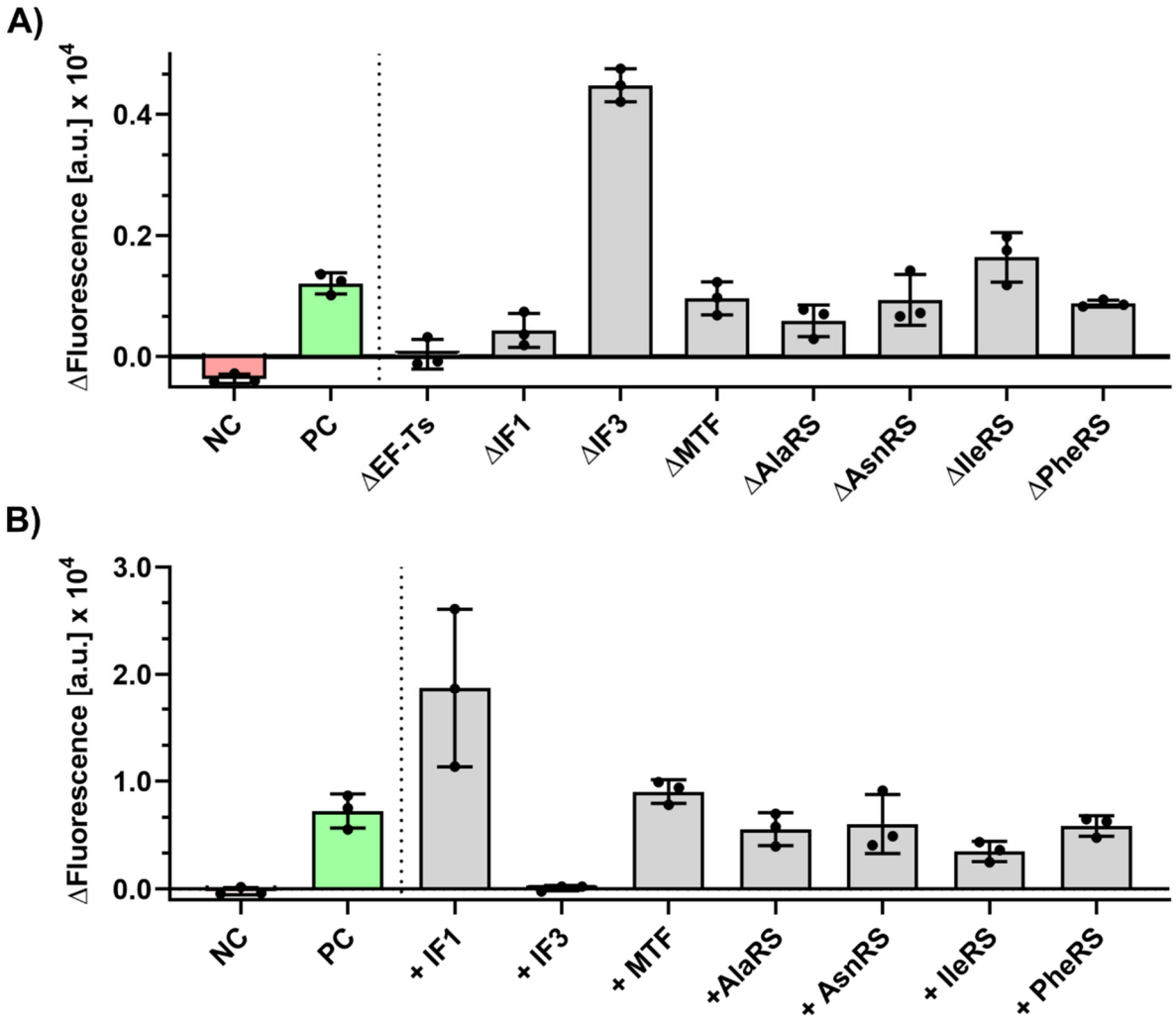


Figure 20 Effect of the individual components of LD2 on replicase activity: A) Endpoint fluorescence for *in vitro* transcription from $[F30\text{-}Bro(-)]_{UTRs(-)}$ in a reduced PURE 3.0 system. Reactions were prepared with 15 µM EF-Tu, 1.5 µM S1, 50 nM RNA, 300 nM MS2rep and the full set of LD2 components (PC, green), or all except the indicated one. The negative control (NC, red) was assembled identical to PC but without the replicase. The reactions were incubated at 37°C for one hour. The LD2 components were all provided at 5 µM final concentration. Error bars indicate standard deviation for three technical replicates. **B)** Endpoint fluorescence for reactions programmed with 50 nM $[F30\text{-}Bro(-)]_{UTRs(-)}$, supplemented with the minimal set of co-factors and the indicated individual LD2 components after incubation at 37°C for one hour. All reactions contained 15 µM EF-Tu and EF-Ts, 1.5 µM S1 and, with exception of the negative control (NC, red), 300 nM MS2rep. The individual LD2 factors were added to 5 µM final concentration. Error bars indicate standard deviation for three technical replicates.

To further investigate these effects, the activity of the minimal replication complex, consisting of the replicase subunit, EF-Tu, EF-Ts and S1, was evaluated in the presence of the individual LD2 components (Figure 20B). In alignment with the previous assumption that IF3 acts as inhibitor, addition of IF3 completely abolished replicase activity. Furthermore, the activity of the core complex, lacking any IF3, was comparable to the PURErred depletion experiment (Figure 19C). Addition of IF1 increased activity by approximately 2.5-fold, IleRS moderately reduced activity and AlaRS did not have a noticeable effect on the activity this time. Except for AlaRS, these results correlate well with the depletion assay, where depletion of IF1 and IleRS lead to a decrease (IF1), or respectively, increase (IleRS). For AsnRS, PheRS, no substantial effect was observed, while addition of MTF potentially increased activity, albeit only to a limited extent.

All seven tested LD2 proteins interact with RNAs under *in vivo* conditions [270–276], thus interactions with the [F30-Bro(-)]$_{UTRs(-)}$ RNA *in vitro* would not be unlikely. Especially MTF and the four aa-tRNA synthetases could potentially interact with the UTR-derived segments, as these can adopt tRNA-like structures [125,277,278]. For IF1, it was shown that it exhibits RNA chaperone activity [279,280], and IF3 was shown to directly interact with MS2 RNA [281–283]. To elaborate, if the observed effects were based on specific interactions between the RNA and the protein, or only unspecific and potentially due to altered crowding/buffering conditions, further experiments were performed.

In a first step, several of the tested proteins were titrated to the MS2rep core complex and the effect on replicase activity evaluated (Figure 21A), with PEG 8000 as a comparison to common crowding agents. In addition, a putative synergy between IF1 and MTF which was observed in preliminary experiments (data not shown), was also further tested. As expected, the effect of IF1 and IF3 had a clear correlation to the concentration of the respective protein factor, where increasing concentrations led to an increase of activity (IF1), or respectively, decrease (IF3). At the highest tested concentration, IF1 (15 µM) enhanced activity to approximately 400%. The highest concentration of IF3 (5 µM) completely abolished activity. Similarly, increasing concentrations of PEG 8000 led to a substantial decrease in activity. In this case, this was presumably due to an increase of dsRNA formation. While MTF alone seemed to provide a moderate increase, no enhancement was observed for increasing concentrations within the range of error. However, in combination with IF1, the effect appeared more prominently, increasing the observed activity to 500%. In case of AlaRS, no distinct effect was seen. In total, these findings further support the hypothesis that IF1 and IF3 act as enhancer, or respectively, inhibitor on MS2 replicase activity.

Before continuing with the characterisation of the *modus operandi* of IF1 and IF3, the replication ability of the [F30-Bro(-)]$_{UTRs(-)}$ RNA was first to be re-evaluated. While *Weise et al.* concluded that for this specific RNA, transcription is only possible from (–) to (+) strand [180], it

seemed reasonable to test this hypothesis under the new conditions including the enhancement by IF1 (Figure 21B).

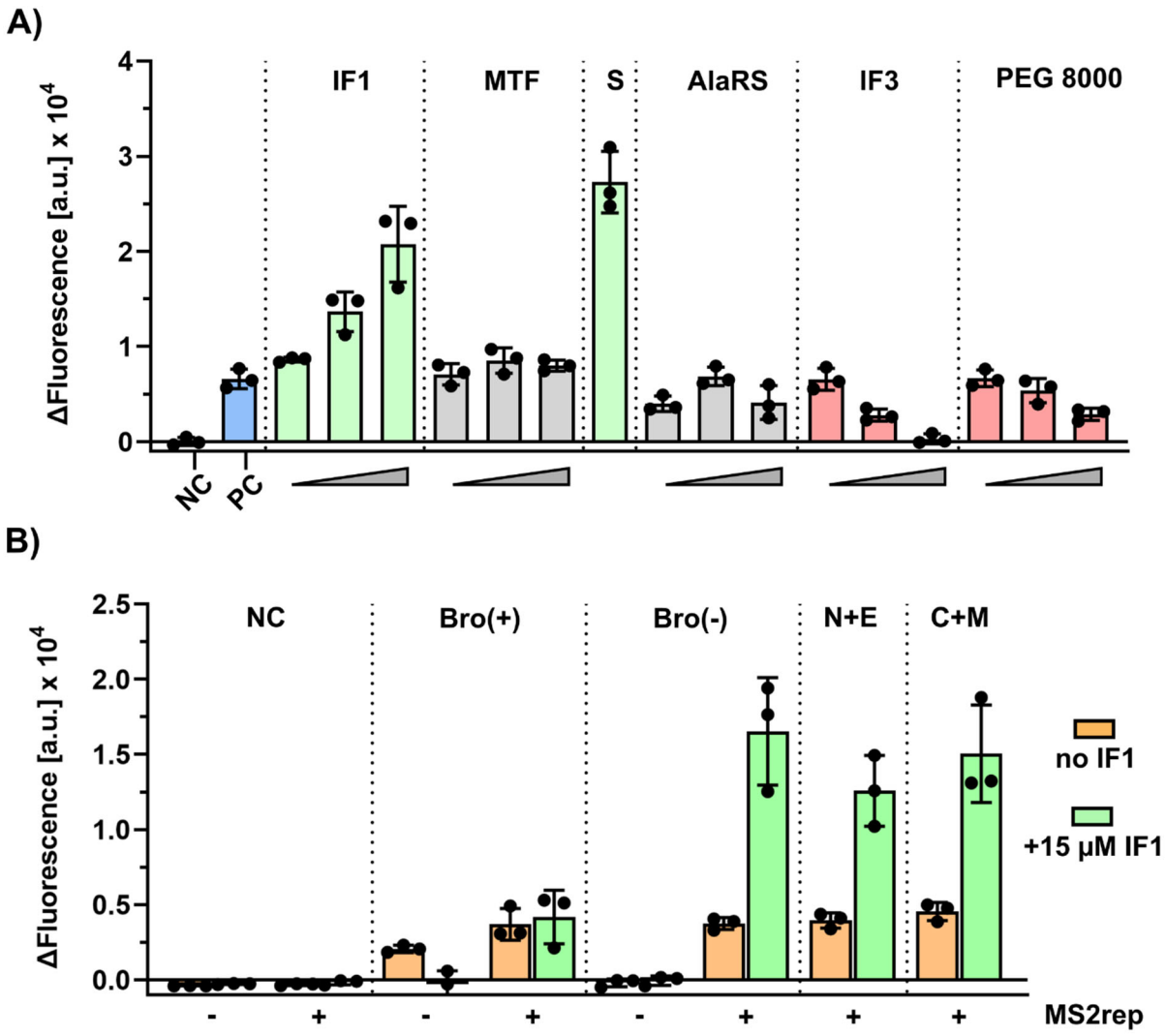


Figure 21 *In vitro* transcription of both [F30-Bro]$_{UTRs}$ strands: A) Endpoint fluorescence of titration experiments for chosen LD2 components and crowding reagent PEG 8000 after incubation at 37°C for one hour. All reactions were assembled with 15 µM EF-Tu and EF-Ts, 1.5 µM S1, 50 nM [F30-Bro(-)]$_{UTRs(-)}$ and, except for the negative control (NC), with 300 nM MS2rep. IF1, MTF and AlaRS were titrated in 1.5 / 5 / 15 µM, IF3 in 0.3 / 1.5 / 5 µM and PEG 800 in 0.3 / 1 / 3 %. Additionally, a putative synergy between IF1 and MTF was tested by supplementing the reaction mix with 15 µM of both (S). Error bars indicate standard deviation for three technical replicates. **B)** Endpoint fluorescence for *in vitro* replication assays of both [F30-Bro]$_{UTRs}$ strands. All reactions contained 15 µM EF-Tu and EF-Ts and 1.5 µM S1. The negative controls (NC) were prepared without RNA, all other reactions were programmed with 50 nM of [F30-Bro(+)]$_{UTRs(+)}$ (Bro(+)) and [F30-Bro(-)]$_{UTRs(-)}$ (Bro(-), RNE, RCE), respectively. MS2rep was added in 300 nM, IF1 in 15 µM. Additionally, a preparation of the same N-terminally tagged MS2rep subunit with EDTA instead of $MgCl_2$ in the purification buffer was tested (N+E), as well as a C-terminally tagged variant with Mg^{2+} in the purification buffer (C+M). The reactions were incubated at 37°C for one hour. Error bars indicate standard deviation for three technical replicates.

Furthermore, two alternative preparations of the MS2 replicase subunit were tested. One where Mg^{2+} was replaced with EDTA in all purification buffers. And one where the His_6-tag was at the C-terminus of the replicase. This was done to determine the conditions, under which the subunit was isolated with highest activity, as in case of Qβrep, substantial effects were observed for these two parameters (Supplementary Figure 9).

For Qβrep, if the subunit was purified in the presence of ETDA, or if it was N-terminally tagged, activity was drastically reduced, even though the reaction buffers were supplemented with sufficient Mg^{2+}. This implies that addition of EDTA, as well as a N-terminal modification, lead to a possibly irreversible alteration of the purified complex, even though no apparent differences were visible with SDS-PAGE (Supplementary Figure 7). For MS2rep, the transcription of $[F30\text{-}Bro(+)]_{UTRs(+)}$ from $[F30\text{-}Bro(-)]_{UTRs(-)}$ was not affected by either of these two parameters. Indeed, considering the range of error, the three tested MS2rep preparations performed identical, with and without IF1 (15 µM) (Figure 21B). This constituted an additional difference between the two replicases.

The observed fluorescence changes for the potential transcription of $[F30\text{-}Bro(-)]_{UTRs(-)}$ from $[F30\text{-}Bro(+)]_{UTRs(+)}$ were not fully conclusive. Even in the absence of MS2rep, a fluorescence increase was reproducibly observed, if no IF1 was supplemented. If IF1 was supplemented, however, fluorescence did not change. When MS2rep was added, an increase was observed for both conditions, although the increase was not substantially bigger than for the control without MS2rep and IF1. At best, this could indicate very low levels of (+) to (–) transcription. In this case, the $[F30\text{-}Bro]_{UTRs}$ construct would be able to undergo the full circle of replication.

More likely though is, that these fluorescence changes represent differences in the folding/binding kinetics of the broccoli aptamer core. The reactions were assembled on ice, before being incubated at 37°C, with a delay between the start of the incubation and the first fluorescence measurement. If IF1 acts as RNA chaperone, it is possible that the broccoli aptamer can fold faster into its functional conformer in presence of IF1 and therefore bind DFHBI-1T more quickly. When IF1 is not present, the folding into the functional conformer and subsequent binding of DFHBI-1T would be slower, and thus result in a bigger difference between the fluorescence at the start and end point. In case of supplementation of MS2rep and IF1, these two might compete for binding to the RNA and delay folding/binding, which would explain why the fluorescence change is bigger in this case compared to respective control. All in all, these findings do not suggest any potential RNA replisome based on the $[F30\text{-}Bro]_{UTRs}$ scaffold, but rather match the findings previously reported [180].

3.6.2 Modulation of replicase activity by IF1 and IF3

In the next step, it was tested if the effects of IF1 and IF3 can also be observed *in vivo* or if they are *in vitro* artefacts (Figure 22A). Therefore, *E. coli* F^+5695 cells were transformed with

expression vectors encoding for IF1, IF3, or AsnRS. Subsequent induction of protein expression was followed by infection and plaque assays. This was done under the assumption that a modulation of replicase activity will also affect the number of pfus and/or plaque morphology. AsnRS served here as a control for the effect of protein overexpression on MS2 infectivity and plaque morphology.

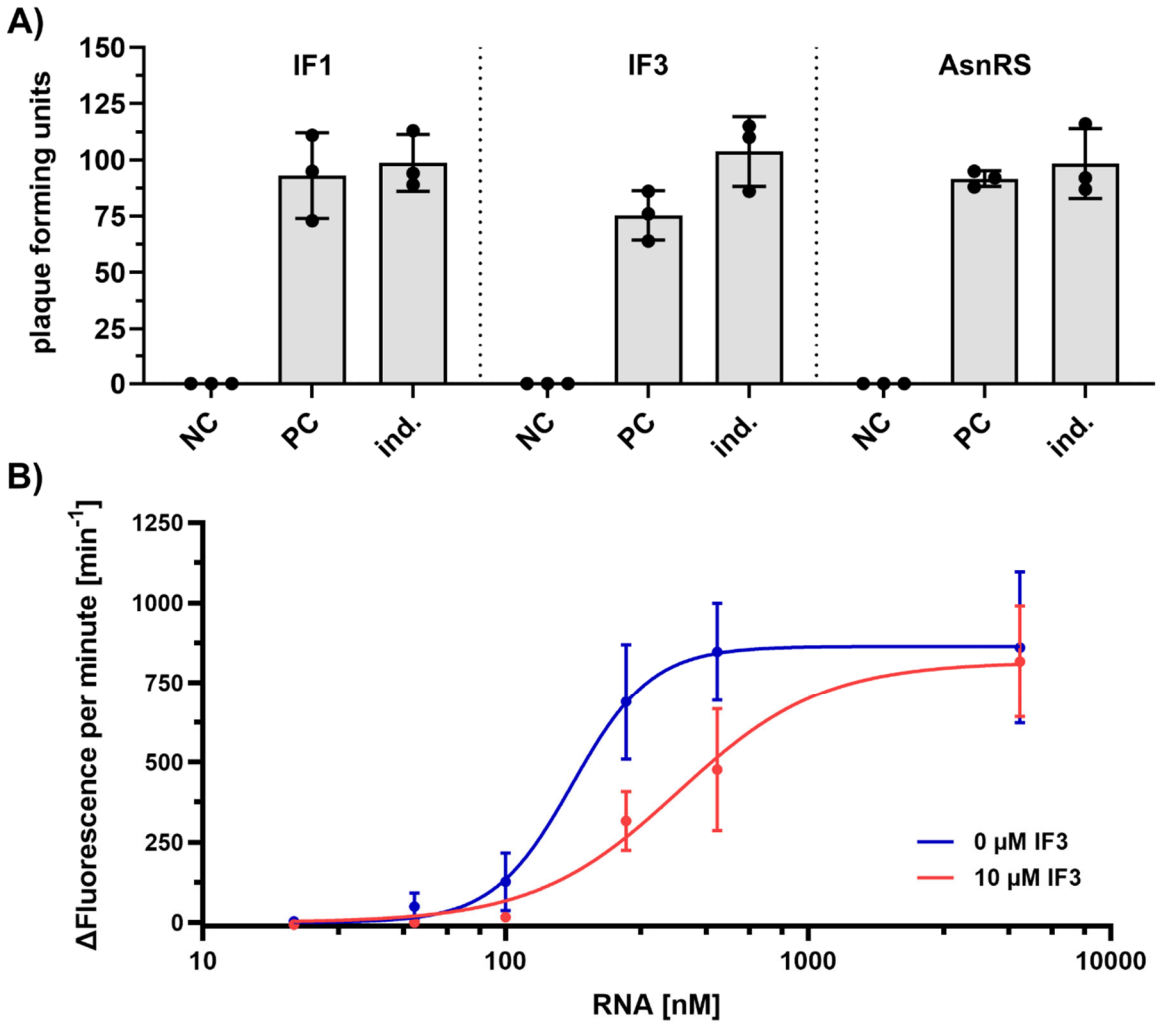


Figure 22 *In vivo* and *in vitro* effect of co-factors on replicase activity: A) Comparison of pfus in dependence of overexpressed IF1, IF3 and AsnRS. *E. coli* F^+5695 cells were mixed with approximately 100 pfu (PC, ind.) or just water (NC) and subjected to plaque assays. To detect potential *in vivo* effects of the chosen LD2 components on MS2 *in vivo* behaviour, cells harbouring the corresponding expression vectors (NC, ind.) were induced with L-Arabinose (0.2 %) one hour before infection. Error bars indicate the standard deviation for three independent, biological replicates. **B)** Initial reaction rates of RNA synthesis by MS2rep (300 nM) in presence (red) or absence (blue) of IF3 for varying concentrations of $[F30\text{-}Bro(-)]_{UTR(-)}$ RNA (20 / 50 / 100 / 250 / 500 / 5000 nM), derived from real time fluorescence measurements using a linear fit for the initial linear phase. IF3 was supplemented in 10 µM. Slopes represent binding isotherms, fitted using the Hill equation $Y=v_{max}*X^n/(K^n + X^n)$. X represents the initial concentration of input RNA, K the apparent concentration at half-saturation, n the Hill coefficient, and v_{max} the maximum rate of fluorescence increase for saturating substrate concentrations. The error bars represent standard deviations from three independent technical replicates.

Using this read-out for infectivity, no apparent effects were detected. Although this experimental setup was considerably simple and might not be suited to detect more subtle effects of IF1 or IF3 overexpression, there are two explanations as to why these effects might in general be difficult to detect. First, IF1 and IF3 are essential proteins that are usually expressed at high levels. It is therefore possible, that overexpression related effects cannot be detected, as the MS2 replication system is already saturated with IF1 and IF3, respectively, under *in vivo* conditions. Alternatively, the interaction between the initiation factors and the MS2 RNA might only occur during a specific stage of the cell cycle/ growth phase and potentially involve multiple factors that modulate the interaction for an unknown purpose. In that case the findings of the *in vitro* experiments could not be directly transferred to the *in vivo* situation.

To further investigate the *in vitro* effect of IF3 on the transcription from $[F30\text{-}Bro(-)]_{UTRs(-)}$, the inhibition was challenged by supplementing the reaction with an excess of MS2rep and S1, respectively (Supplementary Figure 10). While an excess of S1 did not rescue replicase activity, the excess of MS2rep not only restored but even increased the transcription rate. As IF3 can bind to the 3'-UTR of genomic MS2 RNA, it also should bind the homologous 3'-end of the $[F30\text{-}Bro(-)]_{UTRs(-)}$ RNA. There, it potentially blocks transcription initiation by rendering the 3'-end of the RNA inaccessible for the replicase. With an excess of MS2rep, the bound IF3 gets outcompeted and transcription initiation can now take place.

To quantify this inhibition, increasing concentrations of input $[F30\text{-}Bro(-)]_{UTRs(-)}$ were titrated against constant concentration of IF3 and MS2rep and the reaction rates were measured for each condition (Figure 22B). For the non-inhibited reactions, the reaction rates increase until the RNA concentration exceeds MS2rep concentration (300 nM), at which point the replicase subunits present are saturated with substrate. When the transcription was challenged with IF3, no activity was observed for low concentrations of the RNA, but still reached the initial maximum reaction rate at high RNA concentrations. Compared to the non-inhibited reactions, the reaction rates in presence of IF3 increased with a substantial delay. These two observations further support the assumption that MS2rep and IF3 are directly competing for the 3'-end of both the genomic MS2 RNA and the synthetic $[F30\text{-}Bro(-)]_{UTRs(-)}$ RNA.

Combining fluorescence anisotropy and EMSA, it was possible to show and also quantify binding of IF3, as well as IF1 to $[F30\text{-}Bro]_{UTRs(-)}$ (Figure 23). Based on the fluorescence anisotropy experiment, a dissociation constant K_D for the interaction between IF3 and the 3'-end of 3.4 µM (95% CI = 2.33 – 4.97 µM, $R^2 = 0.883$) was determined (Figure 23A). However, this was not possible for IF1. In this case, no binding effect on the fluorescence anisotropy was observed. This can be due to the size difference of IF1 (9.1 kDa) in relation to $[F30\text{-}Bro(+)]_{UTRs(-)}$ (132 kDa) and/or the interaction only being of a transient nature. A band shift assay with IF3 showed a clear retention of $[F30\text{-}Bro(-)]_{UTRs(-)}$ RNA, starting between 2.3 and 3.6 µM IF3, which

is in excellent accordance with the results from the anisotropy experiment (Figure 23B). Interestingly, not only two bands, corresponding to one bound and one unbound state, were observed, but the retention directly correlated with increasing input of IF3. This implies that the RNA can be bound simultaneously by multiple IF3.

Using an identical assay for IF1, no band shift was observed. Instead, the band became less sharp with increasing IF1 concentration, exhibiting a tailing blur (Figure 23C). While this is no clear evidence for a direct RNA-protein interaction, it still hints in that direction. When the buffering system for this band shift assay was changed from Tris/acetic acid to Tris·HCl/boric acid, however, a retention of the RNA was also observable for IF1 (Figure 23D), starting at 1.4 and 2.3 µM IF1.

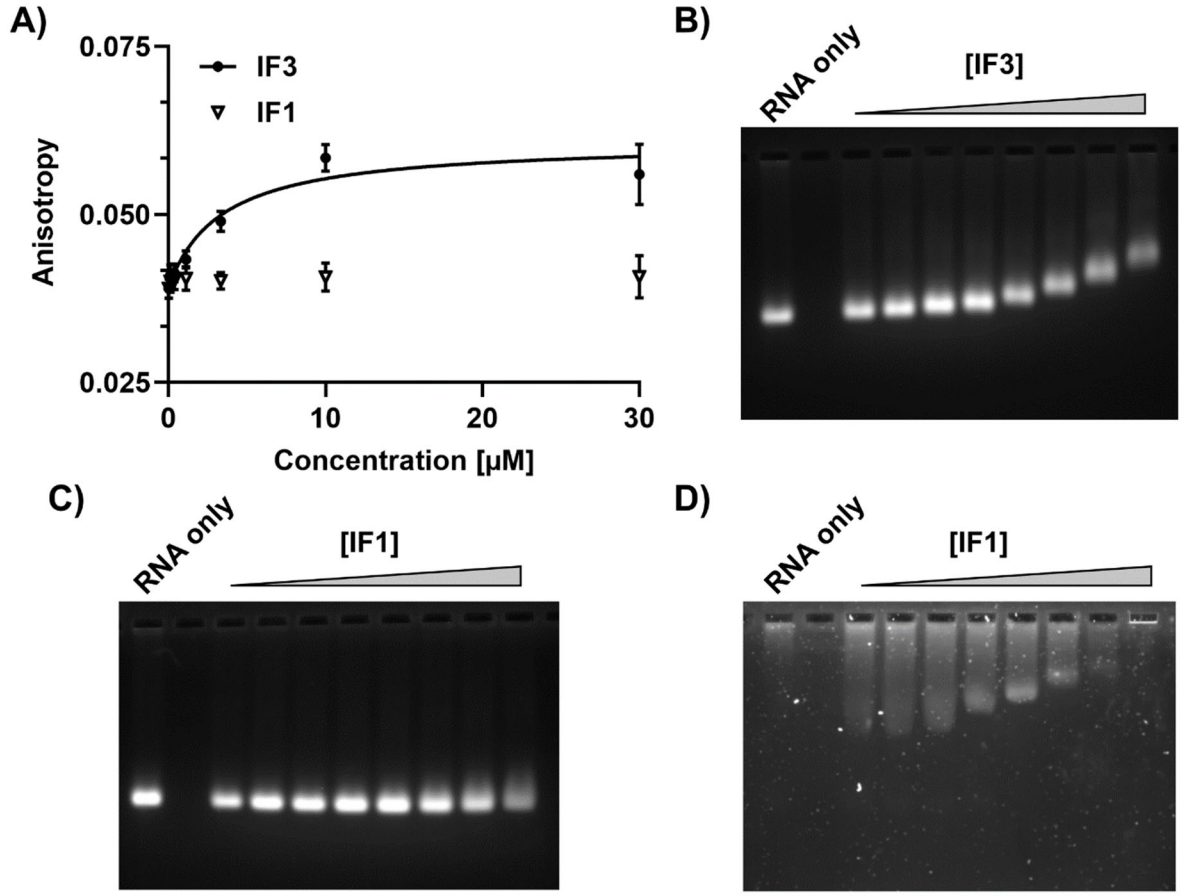


Figure 23 Binding of IF1 and IF3 to [F30-Bro]$_{UTRs}$: A) Concentration dependent effect of IF1 (white triangle) and IF3 (black circle) on fluorescence anisotropy of DFHBI-1T bound by [F30-Bro(+)]$_{UTR(-)}$. DFHBI-1T was provided in 100 nM, [F30-Bro(+)]$_{UTR(-)}$ in 200 nM. IF1 and IF3 were supplemented in 0 / 0.12 / 0.37 / 1.11 / 3.33 / 10 / 30 µM. The following equation was used to fit the data points for IF3, with X as the concentration of IF3, K_D as the dissociation constant, r_{max} the maximum anisotropy at saturating IF3 concentration and r_0 the offset anisotropy: $Y = r_{max}*X/(K_D + X) + r_0$. Error bars indicate standard deviation based on three independent, technical replicates. **B)** and. **C)** EMSA of [F30-Bro(-)]$_{UTR(-)}$ (200 nM) for increasing concentrations (0.56 / 0.89 / 1.43 / 2.29 / 3.66 / 5.86 / 9.38 / 15 µM) of IF3 (**B)**) and IF1 (**C)**), respectively. Samples were loaded onto 1.5 % TA agarose gels (40 mM Tris, 20 mM acetic acid) and run for three hours with 1.5 V cm^{-1} at 4°C. **D)** Same as **C)** but with TB buffer (50 mM Tris-HCl, 50 mM boric acid) as gel and running buffer, instead of TA buffer.

Similar to IF3, multiple levels of retention could be observed, in dependence of the IF1 concentration. This suggests that IF1 also binds in multiple stochiometric ratios to the RNA.

All in all, these findings strongly support the hypothesis that both the enhancing effect of IF1, as well as the inhibiting effect of IF3 on the enzymatic activity of MS2rep are not due to protein-protein interactions but RNA-protein interactions between the initiation factors and the substrate RNA.

3.7 *In vitro replication* of full length MS2

An essential trait for any RNA replisome is that it can be replicated. *In vitro* replication assays can be a good indicator, if an RNA can be used as a replisome. However, there are two criteria that determine the replication capability *in vitro*. First, the tendency to form RNA duplexes. As the RNA should also encode for genes of interests besides the replicase, it must be considerably longer than just a minimal replicator. Consequentially, it might be rendered inactive by a dsRNA formation. To avoid this, the RNA must be able to form stable secondary structures, that will keep the strands separated [266]. Second, the tendency to gradually transform into shorter replicating RNAs [284]. Through a size-depending advantage in replication, these short RNAs, so-called RNA parasites, can potentially outcompete the initial RNA replisome [285–287].

For genomic MS2 RNA, it has been shown that it can be fully replicated *in vitro*, using a fluorescence-based read-out in the PURExpress system, thereby even exceeding the initial input concentrations [100]. The same read-out system was used to elucidate, if the MS2rep core complex is also able to replicate (-) to (+) and vice versa in the minimal buffer system (Figure 24). In brief, the broccoli aptamer was incorporated into the MS2 genome, after the *mp* gene, in both (-) and (+) strand. The corresponding complementary strands were then incubated with the MS2rep core complex *in vitro* and transcription of the respective $[F30\text{-}Bro(+)]_{MS2}$ monitored through fluorescence changes (Figure 24A, B). This process was also evaluated for its dependence on the individual core factors EF-Tu, EF-Ts and S1, as well as enhancer IF1, by depletion, or respectively, addition of the indicated factors (Figure 24C, D).

In both cases, (-) to (+) and (+) to (-) strand synthesis, respectively, was observed in presence of the full MS2rep core complex. Both reactions also heavily depend on both elongation factors, as omission of either one drastically reduced the replicase activity. Interestingly, the activity in absence of S1 was not reduced in comparison to the full core complex. Potentially, the S1 copurified with the replicase subunit is sufficient for the replication for full length MS2 RNA. Alternatively, this process might be less dependant on S1 as expected from the experiments with $[F30\text{-}Bro]_{UTRs}$. These findings were also perfectly reflected by analysis of samples with gel electrophoresis (Supplementary Figure 11).

In contrast, the fluorescence-based assay suggests that IF1 enhanced (-) to (+) transcription,

but (+) to (-) transcription was not substantially affected. However, the in-gel analysis, did not show any difference between both reactions in presence of IF1. This apparent contradiction could possibly be explained by a difference in the stability of the broccoli fold within the MS2 genome. In case of (+) to (-) transcription, the broccoli aptamer would then presumably be less stable, potentially disrupted by base pairing within the MS2 genome. Thus, the observed fluorescence would be lower than expected, based on the amount of synthesised $[F30\text{-}Bro(+)]_{MS2(-)}$ RNA.

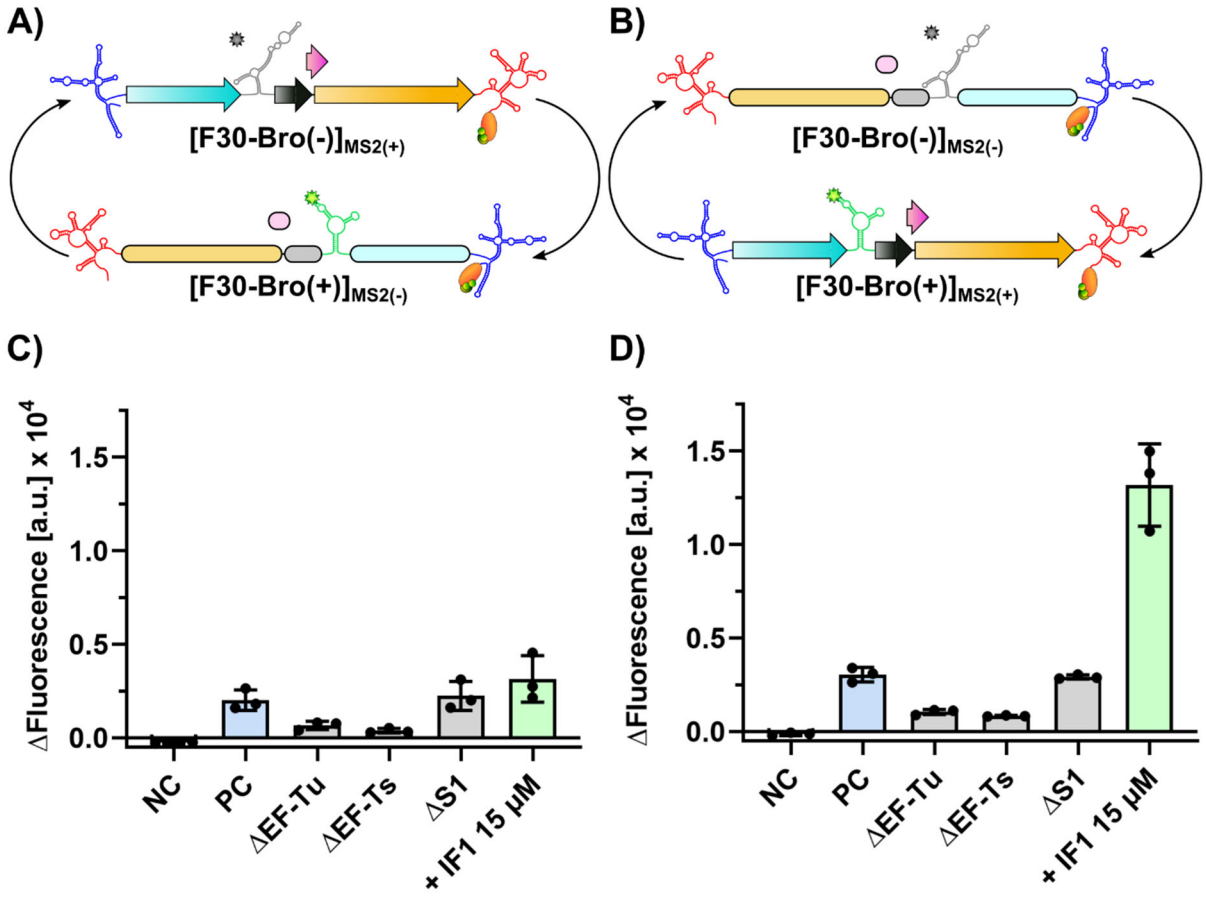


Figure 24 Replication of full length MS2 *in vitro*: A) Scheme for the coupled replication of $[F30\text{-}Bro(-)]_{MS2(+)}$ and $[F30\text{-}Bro(+)]_{MS2(-)}$ by the MS2rep holo-complex. The active, fluorophore binding broccoli aptamer is shown in green, its inactive complementary equivalent in grey. **B)** same as **A)** but for $[F30\text{-}Bro(+)]_{MS2(+)}$ and $[F30\text{-}Bro(-)]_{MS2(-)}$, respectively. **C)** and **D)** Fluorescence endpoint for the *in vitro* replication of the corresponding $[F30\text{-}Bro(-)]_{MS2}$ input RNAs after incubation at 37°C for three hours. The RNA templates were provided in 50 nM concentration and mixed with the essential co-factors as indicated. Both positive (PC) and negative control (NC) contain all co-factors, with EF-Tu and EF-Ts in 15 µM, S1 in 1.5 µM concentration. The NC does not contain MS2rep, while all other reactions were supplemented with 1 µM of the replicase subunit. In addition, 15 µM IF1 as addition to the PC conditions was also tested. Error bars indicate the standard deviation based on three independent technical replicates.

In general, transcription in both directions almost exclusively produced double stranded MS2 RNA, and either small truncated RNAs, or potential RNA parasites. Albeit, it was not evident, if the duplex formation reflected the *in vitro* situation or was an artefact of the handling of the

gel electrophoresis samples. Similarly, no assertion could be made regarding a connection between duplex formation and truncation and/or emergence of short RNAs. A comparison of the protein content of PURExpress and the minimal buffer system might explain the difference regarding the yields of the replication of full length MS2 RNA. The minimal system is lacking ribosomes and further putative RNA binding factors. These however could prevent annealing of genomic and anti-genomic MS2 RNA. By analysing the effect of further single strand binding RNA chaperones on the *in vitro* replication, this hypothesis could be tested.

The replicase of Qβ is notorious to generate short replicating RNAs *in vitro* [284,286,287]. So far, it was unknown if the replicase of MS2 exhibits a similar behaviour. It was also unclear, if the low molecular weight bands observed for full length replication were replicating RNAs. To answer these questions, follow-up *in vitro* replication experiments were performed.

Purified Qβrep was shown to synthesise short RNA replicators even in absence of a designated RNA input [288]. Comparison of untemplated synthesis between Qβrep and MS2rep matched this finding, while also showing that it did not apply to MS2rep under the tested conditions (Figure 25A, B).

The exponential synthesis of short RNAs by Qβrep was preceded by a short delay, and mostly resulted in RNAs between 100 and 200 nt length. The band at around 130 nt likely corresponds to RQ135, the known substrate of Qβrep. Thus, while no actual RNA template was provided, minor contaminations in the lab environment presumably were the source for RNA replication. In case of MS2rep, no bands were detected. This was initially considered as indication that MS2rep is less promiscuous and processive than Qβrep.

In a next step, the *in vitro* behaviour of MS2rep activity was compared over five serial transfers (Figure 25B, C). Therefore, MS2rep was incubated with or without a designated RNA template (genomic MS2), both in absence and presence of IF1 and IF3, respectively. Noteworthy, the experiments in absence of input RNA with no additional IF, as well as both experiments in presence of IF3 were conducted after the other three experiments. The results from these three experiments are in contradiction to the initial assumption that MS2rep was less promiscuous and processive. Thus, once a potent replicator was introduced to the lab during the first round of serial transfer experiments, MS2rep was also capable of considerable synthesis of RNA even in the absence of a designated template (Figure 25C). In consequence, while the overall observations presumably still reflect the expected behaviour for the first serial transfer experiment, the band intensities of the short replicating RNA are most likely overrepresented.

Nonetheless, independent of addition of a designated RNA template, the emergence of short replicating RNAs was observed. For reactions that were supplemented with MS2 RNA, this species was already lost after the initial round (0).

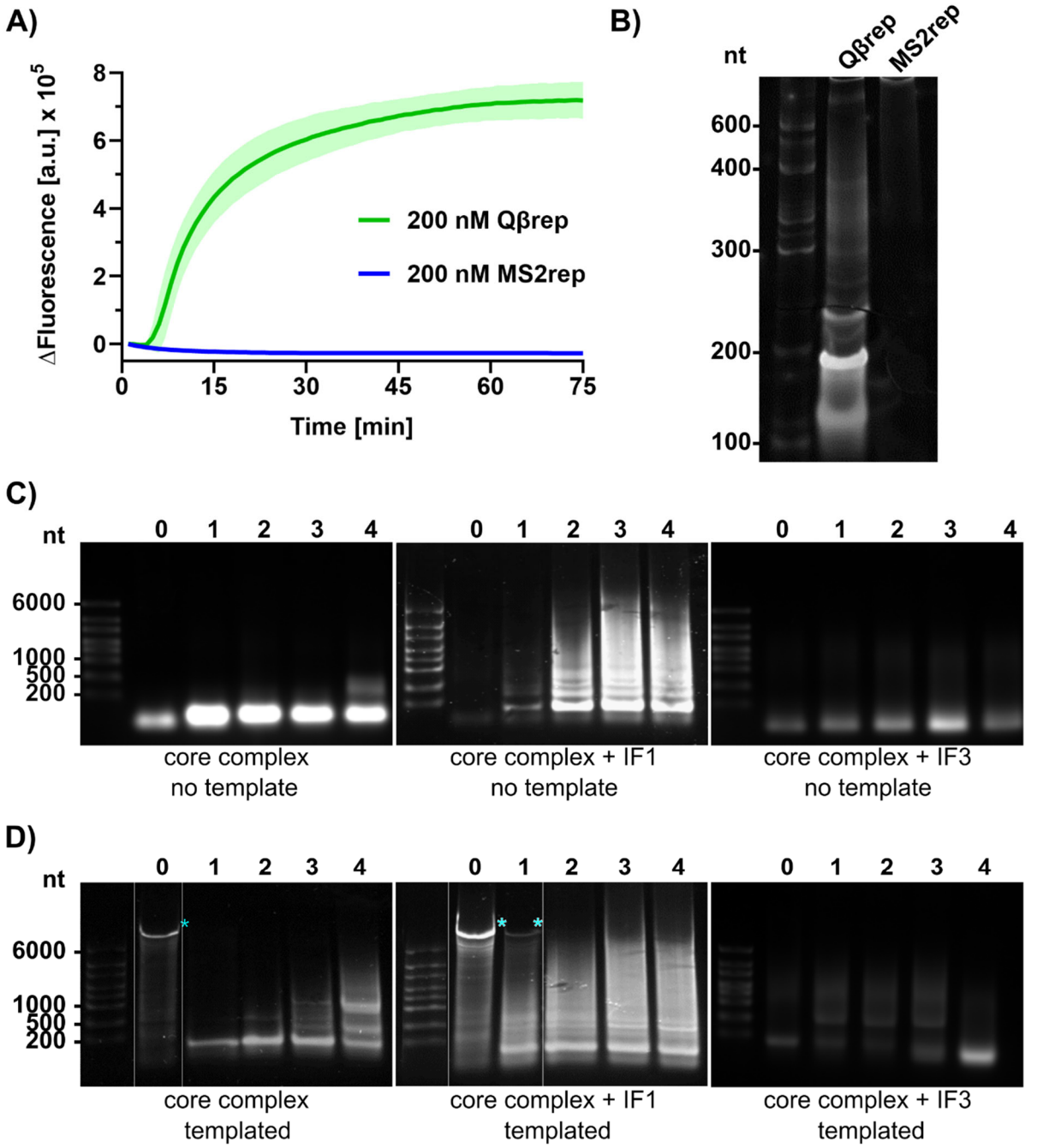


Figure 25 Emergence of small replicating RNAs: A) Fluorescence increase over time in the absence of a distinct input RNA template for reaction mixes supplemented with 15 µM EF-Tu and EF-Ts, 1.5 µM S1 and 200 nM Qβrep, or respectively, MS2rep, incubated at 37°C for 75 minutes. The error bars depict the standard deviation over three independent technical replicates **B)** Gel electrophoresis of samples taken from **A)**, run on a 10 % TBU polyacrylamide gel. The visible bands correspond to small, replicating RNA parasites of Qβrep. **C)** and **D)** Serial transfer experiment for the replication of full length genomic MS2 RNA by MS2rep. Numbers correspond to the round of transfer. The initial reactions were assembled by mixing EF-Ts (15 µ), EF-Tu (15 µM), S1 (1.5 µM) with MS2rep (1 µM, left) and additionally with either 15 µM IF1 (middle) or IF3 (right). Reactions were incubated at 37°C for three hours and used as template for the following series, making up 20 % of the volume of next reaction. Samples were run on 1 % TAE agarose gels. For **D)**, the initial reactions were additionally supplemented with 50 nM of full-length genomic RNA. White vertical lines depict where lanes were moved for better visualization, asterisks indicate double stranded MS2 RNA.

In case of no added IF1 and added IF1, this probably was due to formation of replicative inactive RNA duplexes, which were carried over to the next round. In case of added IF3, the inhibition of RNA replication was the most likely cause for this loss.
Interestingly, addition of IF3 did not show full inhibition of the replication of the emergent RNA parasites. Presumably, the parasitic RNAs can either not be bound by IF3 due to a lack of optimal binding sites, or potential binding sites are inaccessible in the RNA structure. Comparison of reactions with and without IF1 show that, as expected, the presence of IF1 accelerated the emergence and enhanced the synthesis rate of the short RNA replicators.

Most remarkably, the initial emerging parasite is of very short length (~ 200 nt) compared to the MS2 genome (3569 nt), but gradually evolves into longer RNAs (up to 1000 nt). This was unexpected for two reasons. First, it was assumed that the genomic MS2 RNA would gradually decline into shorter RNAs until one or multiple species would reach a stable length. Second, these results are contradicting to the concept of the 'tyranny of the shortest', which predicts that shorter sequences will outcompete longer ones as they are replicated faster [284,289]. This suggests that the first short replicator might be the result of recombination or similar process, following the 'tyranny of the shortest' logic, where the replicator gains a drastic size-dependant advantage over the full-length RNA. Subsequently, the dominant replicator could recombine again, now with residual full-length, fragmented MS2 genome, or itself. This could lead to the emergence of longer RNAs. These secondary parasites might be competitive with the initial parasite. For one, the size-dependant effect during replication might not be as prominent anymore. But they also could have adopted further RNA structures through mutations that provide a selection advantage, for example through facilitation of replication initiation. In conclusion, these new RNA species might serve as a new starting point for the development of a MS2rep based RNA genome.

3.8 Characterisation of the replicating RNA MSRP-22

To elucidate the identity of the novel short replicating RNAs, they were isolated from the serial transfer samples, reverse transcribed and inserted into a cloning vector. Sequencing of multiple clones revealed one sequence of 225 nucleotides length, MSRP-22, sharing high homology with the UTRs of genomic MS2 RNA (Figure 26A, B). Here, the first 118 nucleotides of the MS2 5'-UTR could be fully matched to the 5'-end of MSRP-22. The 3'-end of MSRP-22 could be almost perfectly aligned to the last 105 nucleotides of the MS2 3'-UTR, except for four nucleotides (Figure 26B). In the replication of MS2 RNA, the UTRs play an essential role [180]. MSRP-22 is a recombination of the genomic MS2 UTRs and might thus harbour all elements required for full replication.
Indeed, *in vitro* transcribed MSRP-22 RNA could be specifically amplified by MS2rep (Figure 26C). Furthermore, the synthesis of nascent MSRP-22 correlated with the concentration of

input RNA. In contrast, even at the highest concentration, MS2rep did not exhibit any replication of RQ135. This result emphasizes another difference in the replication machinery between MS2rep and Qβrep. Additionally, it also opens the possibility for future cross and/or orthogonal replication systems based on the two replicase subunits, using the same co-factors.

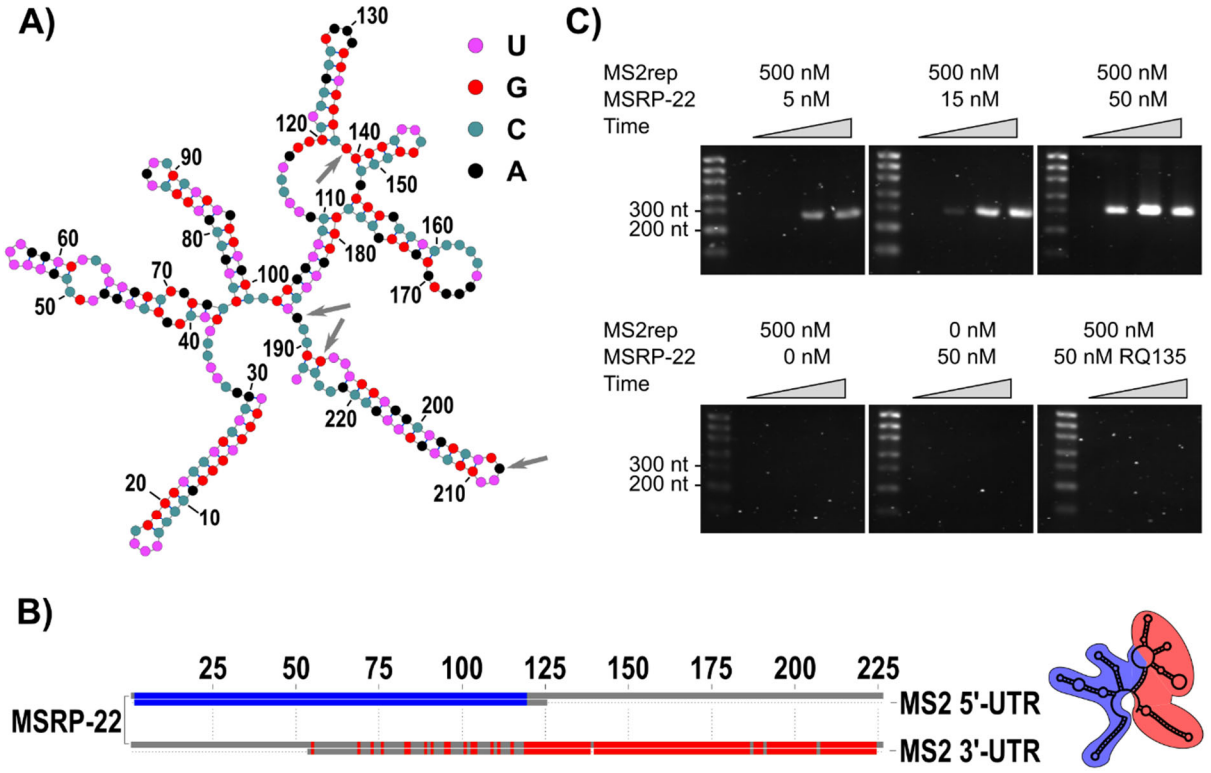


Figure 26 Characterisation of the short replicating RNA MSRP-22: A) Prediction of the minimum free energy structure of the (+)-strand of MSRP-22. The structure was predicted using the web application of RNAfold with default parameters [290,291]. Nucleotide identities are coded by colour: U = pink, G = red, C = petrol, A = black. The grey arrows indicate mismatches regarding the 3'-UTR of the wildtype MS2 RNA used in this study. **B)** Alignment of the 226 nt long MSRP-22 RNA to the first 125 nt of the 5'-UTR (top, blue) and the last 170 nt of the 3'-UTR (bottom, red) of the genomic MS2 RNA, including a depiction of which structural elements correspond to which UTR. Jalview v2.11.1.3 was used for the alignment [292]. **C)** Gel electrophoresis of samples taken from *in vitro* replication experiments of MSRP-22 at four time points (0 / 5 / 10 / 15 minutes). Reaction mixes contained 15 µM EF-Tu and EF-Ts, 1.5 µM, 500 nM MS2rep and were programmed with 5 (top left) / 15 (top middle) / 50 nM (top right) MSRP-22 RNA. Control reactions (bottom) were assembled without RNA (left), MS2rep (middle) or with the Qβ RNA parasite RQ135 (right). The samples were loaded onto 2 % TAE agarose gels.

With a functional MS2 derived replicator, the next question was, if it can encode for additional information and still be replicated. As a starting point, the F30-Bro sequence was incorporated at five different positions, three on the MSRP-22 (+)-strand and two on the (-)-strand. Hereby, nomenclature of (+)- and (-)-strand follows the orientation of the MS2 RNA to which the MSRP-22 strands correspond to (Figure 27A).

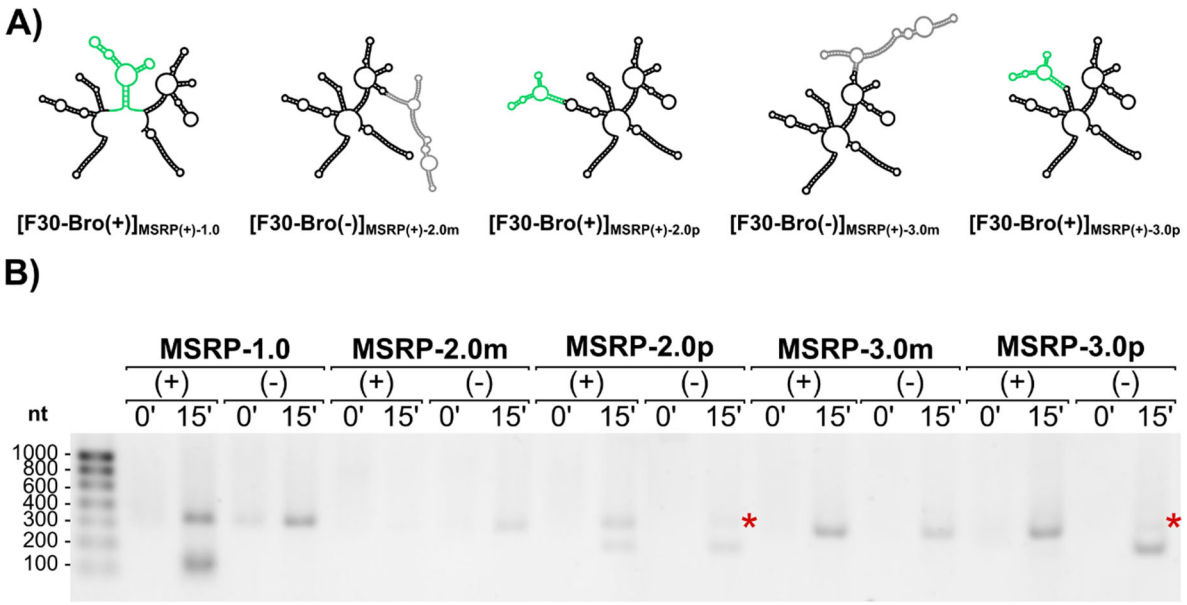


Figure 27 MSRP-22 based fluorescence read-out constructs: A) Depiction of the (+)-strands of five MSRP-22 derivatives, harbouring the F30-Bro module for fluorescence readout. For constructs with the broccoli aptamer on the (+)-strand, the F30-Bro module is shown in green, for constructs with the module on the (-)-strand it is shown in grey. **B)** Samples taken from *in vitro* replication assays of the five [F30-Bro]$_{MSRP}$ constructs shown in A), run on a 1.5 % TAE agarose gel. The samples were taken after 0 and 15 minutes. Reactions were prepared with 15 µM EF-Tu and EF-Ts, 1.5 µM S1 and 250 nM MS2rep and programmed with 100 nM of either (+)- or (-)-strand of the corresponding construct. Red asterisks represent bands of the expected product that were barely visible.

In case of the (+)-strand, the broccoli aptamer was inserted at positions 117 (MSRP-1.0), 52 (MSRP-2.0p) and 80 (MSRP-3.0p). For the (-)-strand, insertions were placed at positions 52 (MSRP-2.0m) and 90 (MSRP-3.0p) of the respective strand. Except for MSRP-1.0, all constructs harbour the insertion in a pre-existing hairpin, while MSRP-1.0 was modified within a loop region. *In vitro* replication assays with both strands of these five constructs showed that indeed at least one strand in all cases could be amplified (Figure 27B, Supplementary Figure 12). In fact, only for [F30-Bro(-)]$_{MSRP(+)-2.0m}$ no apparent amplification was detected. For [F30-Bro(-)]$_{MSRP(-)-2.0p}$ and [F30-Bro(-)]$_{MSRP(-)-3.0p}$, the transcription of the (+)-strand from the (-)-strand was considerably weaker than the opposite direction. In conclusion, four constructs were obtained that were still capable of a full replication cycle and had potential to encode for genes of interest.

To follow up on these promising results, the broccoli aptamer was replaced with the resistance gene against Zeocin (*zeoR*), or respectively, the *rep* gene of MS2, preceded by a ribosome binding site. The F30 stem was maintained, as RNA structure predictions indicated no interference of the coding sequences with the folding of the MSRP-22 scaffold, when the longer insertions were embedded in the F30 stem. While for *zeoR* all five constructs were obtained, it was impossible to get a clone of [*rep*]$_{MSRP-3.0p}$.

Consequentially, these constructs were subjected to *in vitro* replication assays. In accordance with the results for [F30-Bro]$_{MSRP}$, all constructs, except [*zeoR*]$_{MSRP-2.0m}$ and [*rep*]$_{MSRP-2.0m}$, were capable of the full replication (Figure 28A, B). Interestingly, while no substantial double strand formation was detected for the F30-Bro constructs (Figure 27B), in case of *zeoR* and *rep* both already had a drastic tendency for duplex formation. Noteworthy, the [*rep*]$_{MSRP}$ constructs capable of full replication constitute minimal RNA replisomes, as they encode for the essential replication motifs, as well as the required protein factor to perform the replication.

However, even though replication could be shown *in vitro*, it was not clear if the encoded genes could also be functionally expressed. During preliminary experiments with the [F30-Bro]$_{MSRP}$ RNAs, it was also tested if replication can be monitored in real-time through fluorescence measurements. It was expected, that full replication of these RNAs would result in a higher fluorescence signal compared to the non-replicating [F30-Bro(+)]$_{UTRs(+)}$. In contrast to this expectation, the observed fluorescence was drastically lower. The fluorescence intensity depends on the proper fold of the broccoli aptamer.

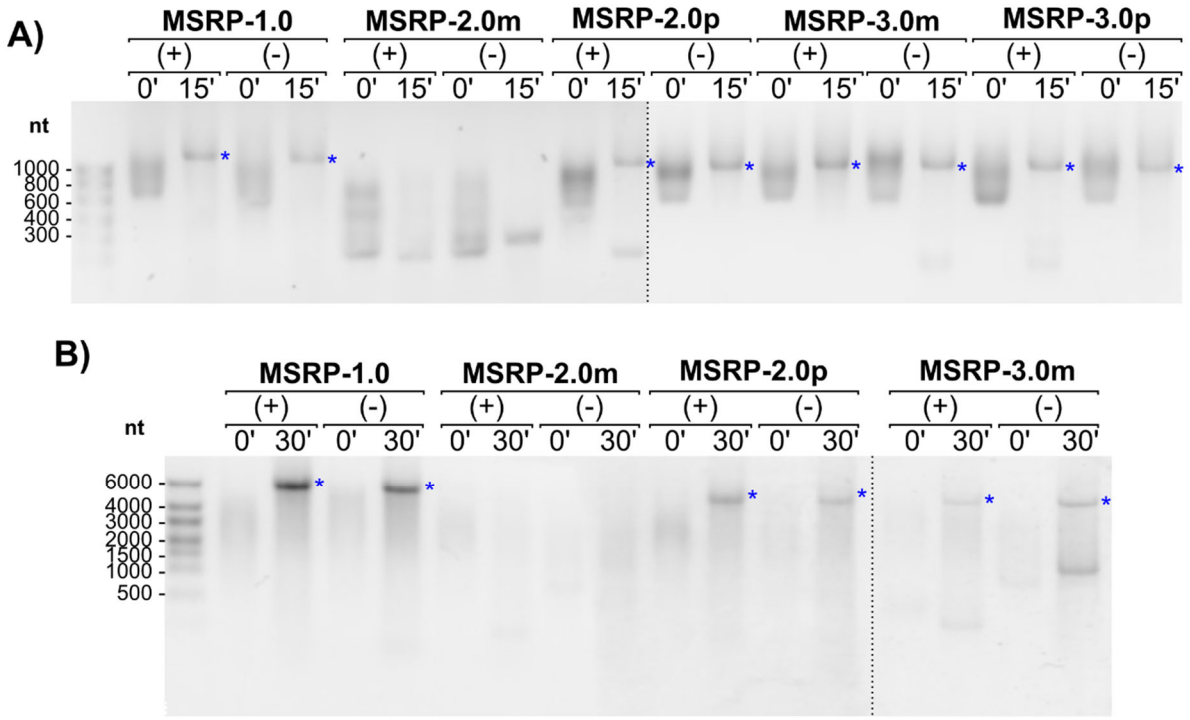


Figure 28 *In vitro* replication of MSRP constructs encoding for genes: A) Samples taken from *in vitro* replication assays of five [*zeoR*]$_{MSRP}$ constructs run on a 1.5 % TAE agarose gel. The samples were taken after 0 and 15 minutes. Reactions were prepared with 15 µM EF-Tu and EF-Ts, 1.5 µM S1 and 250 nM MS2rep and programmed with 100 nM of either (+)- or (-)-strand of the corresponding construct. Blue asterisks represent bands of the double stranded RNA form of [*zeoR*]$_{MSRP}$. **B)** Samples taken from *in vitro* replication assays of four [*rep*]$_{MSRP}$ constructs run on a 1 % TAE agarose gel. The samples were taken after 0 and 30 minutes. Reactions were prepared with 15 µM EF-Tu and EF-Ts, 1.5 µM S1 and 250 nM MS2rep and programmed with 100 nM of either (+)- or (-)-strand of the corresponding construct. Blue asterisks represent bands of the double stranded RNA form of [*rep*]$_{MSRP}$. Black dotted lines highlight where separated lanes were moved together for improved visualization.

Thus, this observation strongly suggests that the aptamer folding might be either impeded by formation of alternative secondary structures or by the replicase subunit preventing the folding and/or dye binding during replication. Transferred to potential replisomes encoding for genes of interest, this means that translation from the (+)-strand could be reduced/inhibited, if the ribosome binding site is inaccessible. Additionally, ribosome binding might not be compatible with the RNA replication and vice versa.

Therefore, an assay for *in vitro* coupled translation and transcription (IVTxT) was used to demonstrate translation of active replicase. Hereby, the individual [*rep*]$_{MSRP}$ RNAs were incubated in PURExpress, utilizing additional [F30-Bro(-)]$_{UTRs(-)}$ RNA for the fluorescence-based read-out (Figure 29A). Input of [*rep*(+)]$_{MSRP(+)-1.0}$ led to the strongest fluorescence increase. This potentially correlates with the observed strongest amplification during the *in vitro* replication assays (Figure 28B). Activity was also substantial for [*rep*(+)]$_{MSRP-3.0m}$, albeit already reduced by approximately 60% compared to [*rep*(+)]$_{MSRP(+)-1.0}$. In case of the non-replicating [*rep*(+)]$_{MSRP(-)-2.0m}$, fluorescence was further decreased, while also lacking the strong initial fluorescence increase. However, it still exhibited a stronger fluorescence increased compared to the fully replicating [*rep*(+)]$_{MSRP(+)-2.0p}$. This discrepancy presumably arises from the competition between translation and replication, as well as the replication/transcription of the two different RNA species within the PURExpress system. In general, these results show that the [*rep*]$_{MSRP}$ RNAs still function as mRNAs for the expression of the *rep* gene. Especially [*rep*]$_{MSRP-1.0}$ showed high potential for future usage as a replisome scaffold.

Furthermore, it was even possible to increase the complexity of this setup, by changing the read-out construct to the replicating [F30-Bro(-)]$_{MSRP(-)-1.0}$ (Figure 29B). In case of [*rep*(+)]$_{MSRP(+)-1.0}$, [*rep*(+)]$_{MSRP(-)-3.0m}$, and [*rep*(+)]$_{MSRP(-)-2.0m}$, the fluorescence was expectedly lower, by approximately 50 % compared to the assay with [F30-Bro(-)]$_{UTRs(-)}$, while the relative difference between the individual [*rep*]$_{MSRP}$ remained mostly unchanged. For [*rep*(+)]$_{MSRP(+)-2.0p}$, no fluorescence change above background level was observed.

Similar to the previous experiment, this might be the result of a non-optimal coordination of all the individual processes taking place, in this case now also burdened with the replication of the read-out RNA. Even though [*rep*(+)]$_{MSRP(+)-2.0p}$ did not exhibit any activity in this experiment, the *in vivo* situation might differ enough for the activity to be sufficient. Therefore, [*rep*(+)]$_{MSRP(+)-1.0}$ and [*rep*(+)]$_{MSRP(+)-2.0p}$ were both used for the design of synthetic MS2-based RNA replisomes, encoding for two genes.

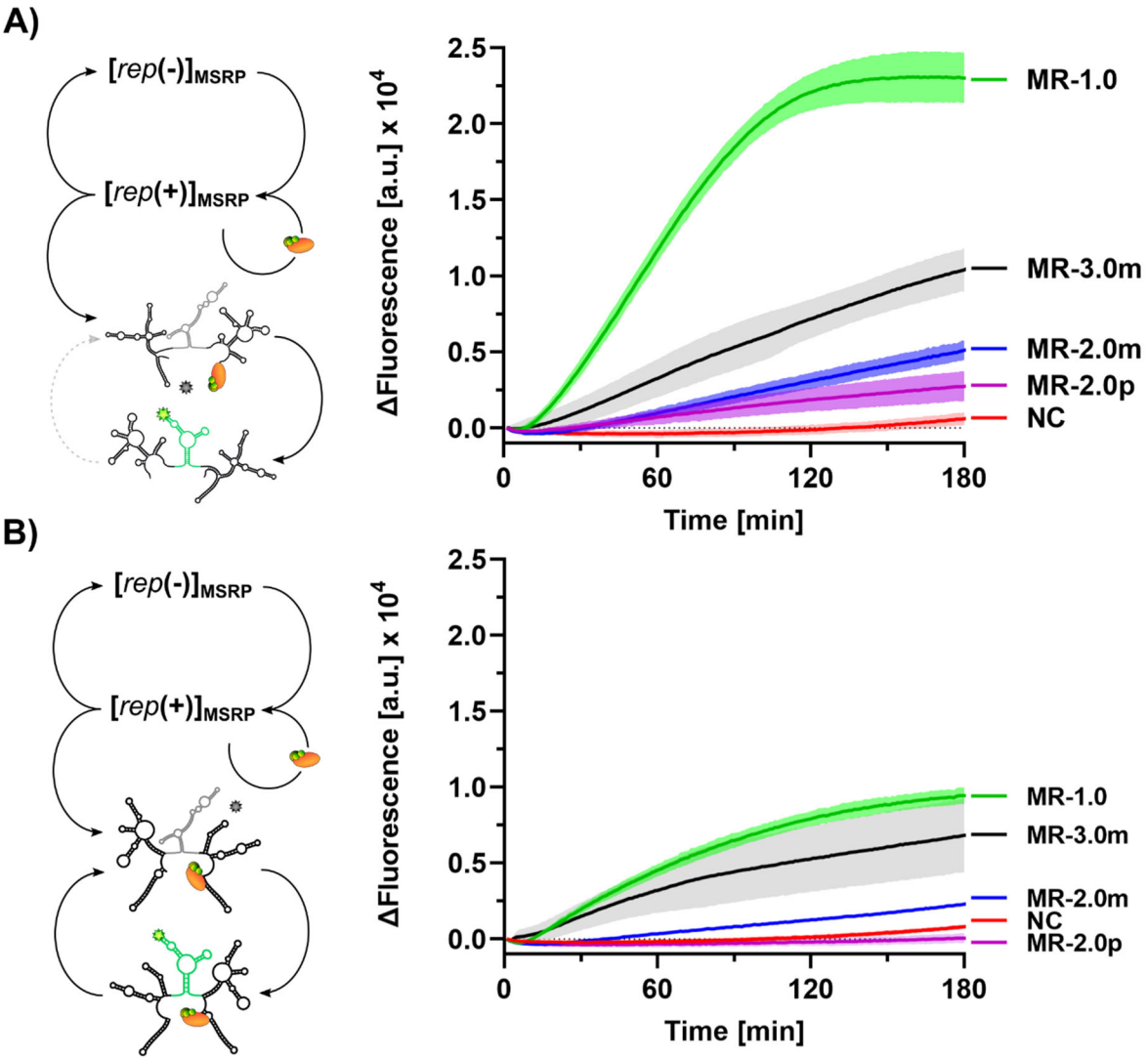


Figure 29 *In cis* and *trans in vitro* replication coupled to *in vitro* translation: A) Right: Scheme of the transcription of [F30-Bro(-)]$_{UTRs(-)}$ into [F30-Bro(+)]$_{UTRs(+)}$ coupled to the *in vitro* translation of the MS2rep subunit from and replication of one of the four [*rep*(+)]$_{MSRP}$ constructs (MR-1.0 / MR-2.0(-) / MR-2.0(+) / MR-3.0(-)). Left: Time course of fluorescence change for this system. Reactions were prepared by programming PURExpress with 100 nM of the corresponding [*rep*]$_{MSRP}$ RNA and 100 nM of [F30-Bro(-)]$_{UTRs(-)}$. For the negative control (NC, red), the [*rep*]$_{MSRP}$ RNA was replaced with water. The reactions were incubated at 37°C for three hours. Error ranges indicate the standards deviation over three independent technical replicates. MR-1.0 (green) = [*rep*(+)]$_{MSRP(+)-1.0m}$ / MR-2.0m (blue) = [*rep*(+)]$_{MSRP(-)-2.0m}$ / MR-2.0p (purple) = [*rep*(+)]$_{MSRP(+)-2.0p}$. **B)** Same as in **A)** but with [F30-Bro(-)]$_{MSRP(-)-1.0}$ as input. In this case, full replication of the readout RNA was expected, based on data shown in Figure 27B.

3.9 Replication of MSRP based RNA genomes

Consequentially, these two replisomes were used to generate the new variants RZ-1, RZ-2, RZ-3, RZ-4, and RZ-5 (Figure 30A). For all five, (+) to (-) transcription and vice versa could be shown, albeit (-) to (+) appeared to be drastically weaker (Figure 30B). Furthermore, only double stranded RNA was detected by in-gel analysis, most likely due to the already mentioned

reasons. In addition, RZ-5 was used as representative for these replisomes for IVTxT together with [F30-Bro(-)]$_{UTRs(-)}$ and [F30-Bro(-)]$_{MSRP-1.0(-)}$, respectively, as read-out template. In comparison to the [*rep*(+)]$_{MSRP(+)-1.0}$ replisome, fluorescence increase, and thus presumably also translation and replication rates were not noticeable affected by the incorpotation of the *zeoR* gene (Supplementary Figure 13).

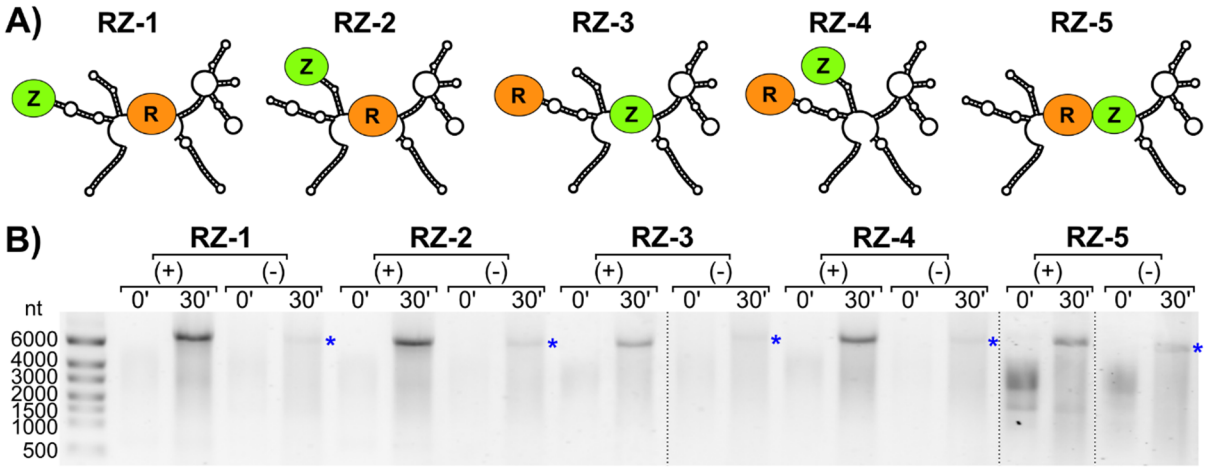


Figure 30 MS2 derived replisomes: A) RNA replisomes RZ-1, RZ-2, RZ-3, RZ-4 and RZ-5, based on the MS2 RNA parasite MSRP-22. The replisomes encode for the MS2 *rep* gene (R, orange), as well as the resistance gene against Zeocin *zeoR* (Z, green). **B)** Analysis by gel electrophoresis of samples from *in vitro* replication assays with both (+)- and (-)-strands of the respective replisome after 0 and 30 minutes. Black lines indicate where separated lanes were moved together to improve visualization. Blue asterisks highlight barely visible bands from (-) to (+) transcription

The next step was to electroporate these replisomes into *E. coli* F$^+$5695 cells and evaluate if they alone can provide the transformed cells with Zeocin resistance *in vivo*. In addition, F$^+$5695 cells harbouring an expression vector for the MS2rep subunit and S1, respectively, were also used. The rational for this was that pre-expressed replicase might help to jump start replication and thus increase replisome levels for sufficient gene expression and replisome persistance. S1 then serves a control, to compare potential differences in *in vivo* replication. In this case, cells were also transformed with the replicating [*zeoR*]$_{MSRP}$ constructs, with the expectation that the over pre-expressed replicase should also be able to replicate these RNAs *in vivo*.

In initial trials, however, no substantial increase of Zeocin resistant cells was observed (data not shown). Though, colony PCR of cells that were transformed with RNA indicated that the resistance in these cases did at least not arise from plasmid contaminations. Albeit, even for the negative controls, where only water was 'transformed', a low number of cfus was observed. This was not unexpected, as the concentration of Zeocin in these experiments was comparably low (12.5 µg/mL) compared to commonly used concentrations (25 – 50 µg/mL).

Nonetheless, two interesting observations were made. First, for cells transformed with

RNA, the colonies in general seemingly exhibited a larger diameter as compared to the negative control. Therefore, while potentially not capable of stable maintenance *in vivo*, they might still have acted as transient mRNA for the expression of Zeocin resistance. Second, for controls where the plasmids encoding for the respective RNA replisomes were transformed, the number of cfus was drastically lower, by probably several orders of magnitude, as would be usually expected from transformations of purified plasmids with a highly expressed resistance gene (see section 3.4). However, the numbers of cfus for transformations of plasmids were still drastically higher than for transformations of RNA. As the plasmids do not have a dedicated promoter for the expression of the RNAs, the observed cfus presumably correspond to background levels of transcription of these RNAs and in consequence background Zeocin resistance. Interestingly, numbers of cfus were reproducibly and substantially lower for plasmids that did only encode for the *zeoR* gene, as compared to the plasmids encoding both *rep* and *zeoR* gene. Both replicator types were encoded in identical backbones (pUCIDT), containing an identical set of functional elements (Ampicillin resistance gene with promoter / ColE1 origin of replication / *lac* operon). In consequence, the most likely source for the observed difference must be the *rep* gene. This could indicate that the low-level background transcription from these plasmids is sufficient to enable synthesis of replication capable RNAs and ensuing translation and replication of them *in vivo*.

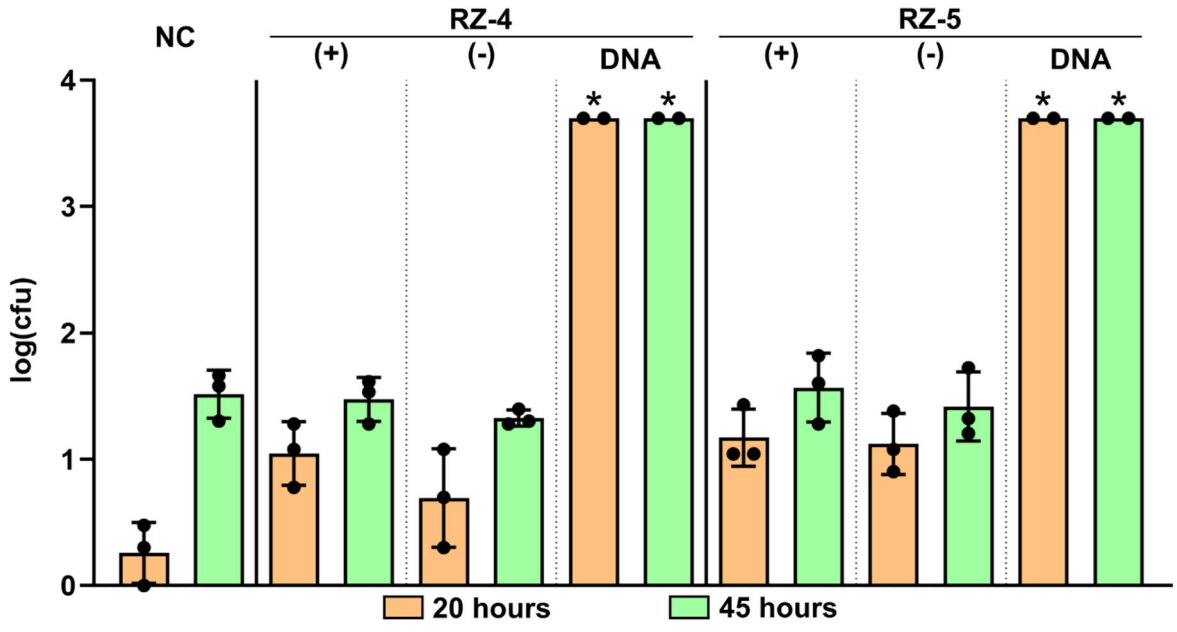


Figure 31 *In vivo* effect of RNA replisomes on Zeocin resistance: Quantification of cfu for cells electroporated with 2.5 µg of (+)- and (-)-strand RNAs, as well as 25 ng of the respective plasmids (DNA) of RZ-4 and RZ-5, after 20 hours (orange) and 45 hours (green) incubation at 37°C, in presence of 10 µg/mL Zeocin. As negative control, cells were transformed with water. Asterisks indicate samples where cfu density was too high to reliable count cfu, and therefore were estimated to be approximately 5000. As this was the case for two replicates for the transformation of DNA, no third replicate was done. Error bars indicate standard deviation from three independent biological replicates.

To quantify these observations, the experiments were repeated with adjusted protocols.

First, for the quantification of a putative function as transient mRNA of the RNA replisomes, cells were transformed with both strands of RZ-4 and RZ-5 (Figure 31). The corresponding plasmids were transformed as positive control, water as additional negative control. In case, that limited antibiotic resistance would slow down cell growth, plates were incubated for 45 hours at 37°C, to allow small colonies to grow to visible size. In addition, per transformed construct, eight colonies were subjected to colony PCR (Supplementary Figure 14).

In general, transformations with DNA led to drastically higher transformation efficiencies compared to transformations with RNA, which was expected. As colony PCR showed, only when DNA was transformed, a band corresponding to the correct amplicon was detectable (Supplementary Figure 14). Thus, the antibiotic resistance of the cells transformed with RNA did at least not stem from plasmid contaminations. Furthermore, cfu counts were higher after 45 hours. However, for counts after 45 hours, no apparent difference between the negative controls ((-)-strand / water) and the actual samples ((+)-strand) was observed. In contrast, after 20 hours incubation, less cfu were visible when cells were only transformed with water, than when RNA was used. The number of cfus in this case were only in the lower two-digit range, even for the additional negative controls where (-)-strand RNA was transformed. Together, these findings suggest that the transformation of RNAs did not result in a measurable Zeocin resistance, and the observed cfu represent only the background noice of cells with antibiotic immunity.

The next step was to quantify the difference between plasmids encoding replicators with and without the *rep* gene (Figure 32). Therefore, cfu counts were compared for cells transformed with the plasmids encoding for [*zeoR*]$_{\text{MSRP-1.0}}$, RZ-1, RZ-3, RZ-4, and RZ-5 (25 ng). To increase the stringency of the ensuing selection with Zeocin, the concentration of it was increased to 37.5 µg/mL. This was sufficient to completely prevent bacterial growth when no DNA was transformed. In contrast, bacterial colonies could grow when plasmids were transformed. Similar to the initial trial experiments, more colonies were observed when the transformed plasmid also encoded for the *rep* gene. In case of the plasmid encoding for RZ-1 and RZ-5, a tenfold increase in cfus was obtained compared to the plasmid encoding [*zeoR*]$_{\text{MSRP-1.0}}$. For the other three plasmids, the number of cfus was approximately 1.5 – 2 orders of magnitude higher in comparison to Z-1.

In conclusion, these findings are promising for future experiments as they suggest that replication and translation of the synthetic RNA genomes can actually happen *in vivo*. Cloning the coding sequences into a plasmid that can get removed from the cells, for example by incubation at elevated temperatures [293], would allow to provide the transformed cells with an initially high level of replicators and antibiotic resistance. The curing of the plasmid then

enables the removal of the DNA source of these RNAs from the cells at a later stage. In the process of plasmid removal, selection pressure would favour cells that harbour RNA genomes successfully adapted to the *in vivo* conditions. In consequence, improved replisomes could be isolated and characterised, eventually enabling a broader variety of encoded functions.

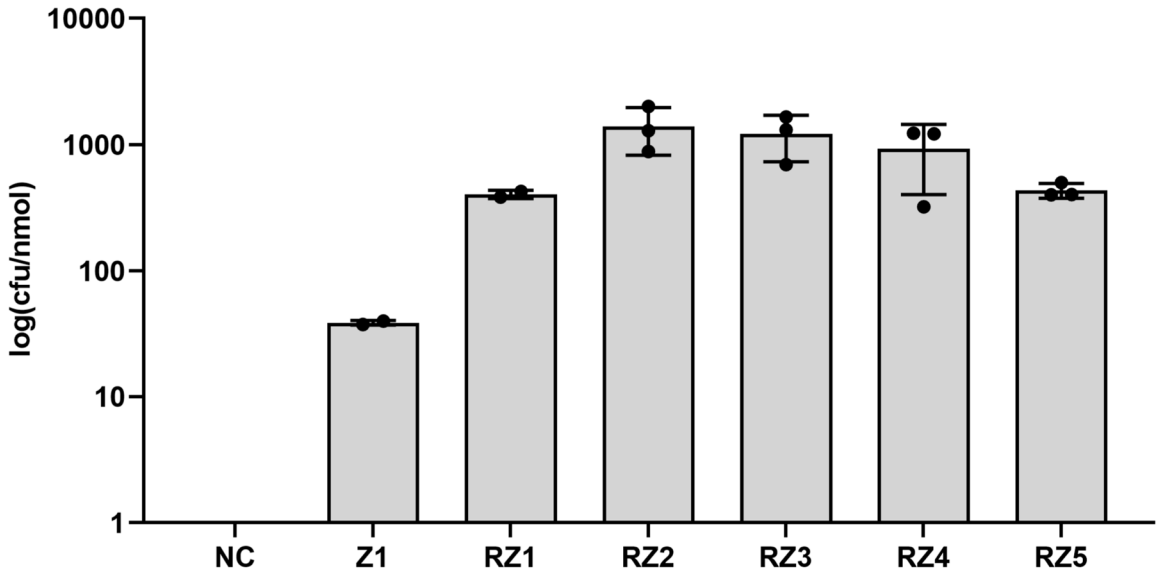


Figure 32 Comparison of Zeocin resistance from transformed plasmids: *E. coli* F^+5695 were transformed with ~7.8 fmol of the plasmids encoding for RZ-1, RZ-2, RZ-3, RZ-4, and RZ-5, and ~12 fmol of pUCIDT_MSRP-1.0(zeoR) (Z-1), respectively, incubated for one hour and then $1/10^{th}$ of transformed cultures was plated on 1.5 % LB-agar plates with 37.5 µg/mL Zeocin. Cells were transformed with water as negative control (NC). Counted numbers of cfus were normalised to cfus per fmol input. Error bars reflect the standard deviation from three independent biological replicates, except for Z-1 and RZ-1 where for each plasmid one outlier was ignored.

4. Discussion

4.1 What might happen to the maturation protein *in vivo* and why

Reports from pioneering experiments on the maturation protein of MS2 are contradicting regarding its fate after infection. Findings by *Krahn et al.* suggest that cleavage of the maturation protein occurs upon or subsequently to binding to the F-Pili, potentially by proteases in the periplasm [294]. In contrast, earlier reports by *Kozak et al.* and *Zinder* et *al.*, suggest recycling of parental maturation protein in the progeny phages [114,295].

Here, it was demonstrated that the MP was sensitive to degradation starting from the N-terminus. Therefore, the observed fragments after infection [294] could be intermediary products from this digest. The removal of the MP from the RNA is required for the initiation of the translation of the MS2 genes and subsequent RNA replication [295]. Proteolysis might here accelerate this removal or prevent rebinding after dissociation. In addition, the formation of the mature MS2 virion depends on cooperative binding of coat protein and MP to the RNA [106,118,130,296,297]. The loss of the coat protein shell upon RNA internalization might destabilize the MP-RNA complex which in turn would more easily dissociate inside the infected cell. This would further accelerate removal of the MP and therefore lead to earlier RNA replication.

It is speculative however, if and how conformational changes within the MP and/or the RNA that surely accompany this dissociation, and degradation of the MP are connected. Recent findings by *Dai et al.*, suggest that the interaction with the MP inside the mature MS2 capsid enforces structural changes of the respective bound RNA hairpin [105]. Like a spring that is compressed, this conformation could store energy. As the MP-RNA complex is stable enough to be infectious however, the release of this energy must somehow be locked. Degradation from the N-terminus might act as a trigger to release this energy, driving the release of the RNA. The observed recycling of MPs might then arise from proteins that in some way evaded degradation and/or were somehow trapped on the RNA and could either rebind RNA or directly be repackaged. Interestingly, findings presented in this thesis highlight the important role of the destabilizing amino acid F4 in MS2 infection, potentially at the level of RNA binding. This supports the proposed hypothesis for RNA release.

However, further research requires optimization of the current purification protocol, especially regarding the removal of nucleic acid contaminations. An RNA-free MP would allow for more elaborate binding studies with MS2 RNA, and even structural analysis of the MP-RNA complex. Furthermore, findings from follow-up research might enable the design of a reliable MP-based delivery system for RNA.

4.2 Expanding the toolbox

4.2.1 The reductive approach to reprogramming 2.0

The MuA transposition system proved to be a highly efficient approach for the generation of large mutant libraries. While the MS2 mutants generated in this study did not exhibit any detectable activity, when tested *in vivo*, several reasons can be found that explain this. For one, formation of long double-stranded stretches on the RNA could prevent MS2rep from replicating the RNA. However, this can potentially be circumvented by switching to MuA recognition sites with lower complementary or modifying the sequences of the current recognition sites.

Second, the high background of mutants that carry the insert outside the desired region effectively reduces the number of putative active mutants. In contrast to the protocol for the generation of the random insertion libraries established for this study, it is also possible to use the MuA transposition system on linear DNA [265], for example PCR amplicons of the desired regions. Following this approach, the potential for off-side insertion is completely removed. The thus obtained modified amplicons can then be ligated into an empty cloning vector and efficiently selected for.

And third, the complex coordination of RNA replication and translation of the selection marker can reduce the chances of identifying active replisomes. Thus, the expression of the selection marker can be below a required level for the respective antibiotic concentration, or replication can be too slow to enable maintenance of the replisome throughout cell division. In consequence, promising transformants will either be killed off before they can develop full resistance, or they cannot be discriminated from background growth of colonies. To enhance the selection advantage of transformants with a functional replisome, the replicase can either be pre-emptively expressed in the cells or the replisome separated into smaller fragments, thus decrease the competition between replication and translation by reducing the number of ribosomal binding sites per strand.

In conclusion, this system might drastically facilitate the generation of novel replisomes in future research, independent of the starting viral vector. This is due to the mechanism of insertion being independent of the target sequence, thus also being applicable for viral vectors that infect eukaryotes. Therefore, new approaches in medical and infection research are possible.

While the screen for MP-binding RNAs was abandoned, the general method of mutagenic IVT still yields significant advantages for future experiments and spin-offs thereof. In contrast to error prone PCR, mutations that appear early on are not propagated and overrepresented, as nascent mutated RNA strands do not serve as template for further transcription. Thus, combination of error prone PCR with mutagenic IVT could substantially improve sequence diversity, and could even allow incorporation of non-canonical nucleotides at the RNA level. In

general, this might be a suitable strategy to generate initial libraries for *in vitro* evolution of the MSRP-22 based replisomes, as well as mutants generated by random insertion with MuA.

4.2.2 Increasing EcI5 promoted disruption efficiency

The disruption efficiencies of the *lacZ* gene by EcI5 introns and the calculated scores for the respective insertion sites, reported by *Zhuang et al.,* did not correlate for all tested target sites [195]. As the algorithm for the computation of these scores only uses nucleotide frequencies derived from insertion into target DNA, other important factors that impact successful disruption are not considered. In the respective section of this thesis, two opposing splicing processes were proposed to also affect disruption efficiency: Reverse splicing of the intron lariat into mRNA and splicing out of the intron from mRNA transcribed from the target locus. Based on which of these two is dominant, the observed gene disruption might drastically deviate from the disruption solely based in successful intron insertion.

To reduce the severity of these effects on the read-out, several approaches are possible. First, limiting the activity of the IEP, for example by using a thermosensitive variant that can be inactivated by elevated temperature or by curing the IEP encoding plasmid, could already reduce the intron activity below a relevant level.

Second, the prediction of secondary structures of the pre-spliced mRNA might facilitate the identification of insertion sites that result in a misfolded and thus less active intron. Longer 5'-exons increase the likelihood of alternative base pairing with the intron part simply by providing more potential interaction partners. Thus, especially insertions towards the 3'-end of a coding sequence might result in a misfolded pre-spliced mRNA. However, in this case functional proteins could still be translated.

Last, choosing target sites on the antisense strand would allow to completely avoid both splicing out from and into the mRNA. In this case, no mRNA harbouring an active intron could get transcribed, nor would wildtype mRNA serve as target for intron transcribed from the plasmid. This approach though requires that a suitable target site can be identified.

The algorithm used for this thesis allows the easy identification of possible insertion sites on both sense and antisense strand. While it has not been considered for this study, the recently discovered additional single nucleotide interaction between the intron (EBS2a) and the target DNA (IBS2a) can increase the number of potential insertion sites fourfold [215]. In the publication by *Zhuang et al*, IBS2a corresponds to the adenine at position -14, which appeared to be essential, but actually is variable, if EBS2a is changed accordingly [195,215]. The algorithm can easily be adapted to also consider this nucleotide as variable position, by adjusting the frequency ratios at position -14. This was also done for other positions to reduce the trainings bias under loose restrictions (see section 2.6.2).

Furthermore, the full algorithm can also identify potential target sites in alternative coding

sequences that use synonymous codons as the input sequence. This in combination with secondary structure prediction and knowledge of the respective enzymatic mechanisms, respectively, could also be used to find targets at specific positions, for example in highly structured regions, or respectively, upstream of catalytically important epitopes. Alternatively, this method could also be used to identify synonymous coding sequences that adopt a predefined fold and rank these by their calculated scores for insertion. While this might in general not be feasible though for genomic loci, it might be beneficial for applications that rely on synthetic reporter genes, for which the codons can be more freely chosen.
In conclusion, these approaches could improve the correlation between insertion of the intron and disruption to the target gene. Finally, this will help to adapt the algorithm for computation of scores with higher validity regarding the disruption of target genes.

4.2.3 Increasing EcI5 insertion efficiency for RNA only approaches

The group II intron EcI5 is of special curiosity for applications in combination with self-replicating RNAs. In general, it has high potential for gene engineering, as a diverse number of administration approaches is possible (Figure 33A). While in this study, the intron/IEP system was successfully expressed as a single RNA from an inducible plasmid, two more alternatives are possible. On one side, this single mRNA can directly be delivered into cells, similar to Cas9-mRNAs in genome engineering [298,299]. On the other side, as is already being practiced for CRISPR/Cas systems, an *in vitro* pre-assembled RNP consisting of the intron RNA and the EcI5 IEP [300] can also directly be delivered to target cells [301–303]. Both approaches have similar advantages over the usage of plasmid encoded EcI5 intron. Thus, they do not require the curing of the plasmid after the gene engineering, but also reduce the risk of accidently generating a mobile genetic element and are only transiently active. These putative applications themselves pose substantial reason for further research on the EcI5 intron. In addition, the intron, which in nature already encodes for the IEP, could also be modified to encode for any gene of interest. In combination with the DNA free methods of delivery, the intron/IEP system could be harnessed as a potent programmable gene editing tool.
However, the combination of this intron with an RNA replisome widens the range of novel applications even more. Thus, embedding the intron within a self-replicating RNA has the potential to amplify the intron RNA *in vivo*, thereby increasing the chances of successful gene engineering (Figure 33B). If replicase and IEP/intron are encoded on opposing strands of a single replisome, expression of the replicase from the $(+)_{rep}$ strand would be followed by synthesis of the corresponding $(-)_{rep}$ strand. The nascent $(-)_{rep}$ strand serves as mRNA for the translation of the IEP, but also contains the active intron RNA. Self-splicing of the intron leads to the formation of the intron lariat, as well as a shortened $(-)_{rep}$ strand, now only encoding for the IEP and the complementary CDS of the replicase.

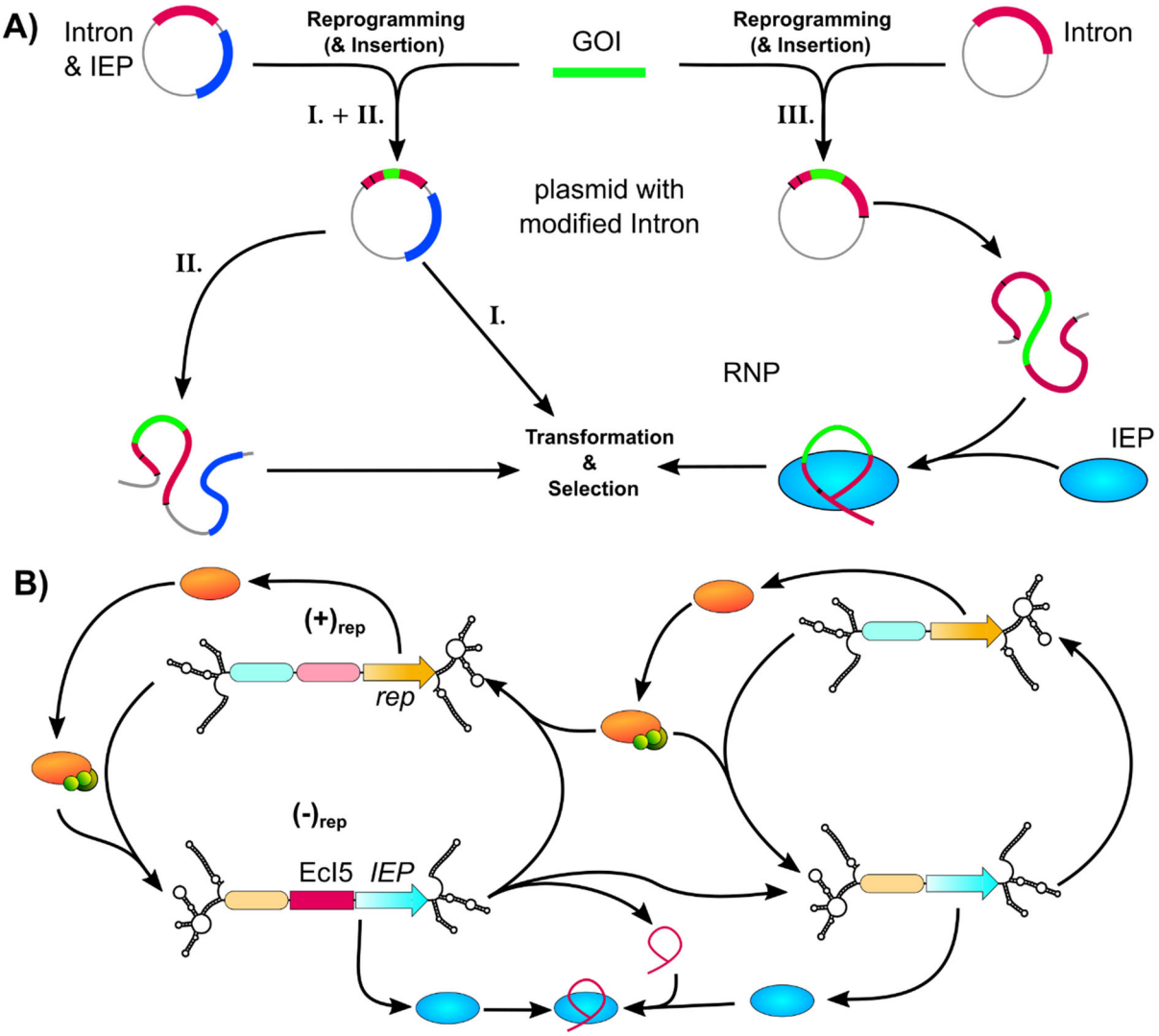


Figure 33 Gene engineering with EcI5 intron: A) Scheme for the gene engineering using the EcI5(ΔORF) intron. (I) The plasmid embedded EcI5 intron construct, which encodes for a 'bicistronic' mRNA (intron & IEP) can be reprogrammed to any suited insertion site, while also optionally being modified to harbour any cargo gene (GOI). The modified plasmid can subsequently be transformed into cells that carry a system for induction of transcription/translation of the EcI5 RNA. This was the approach used in this study. (II) Alternatively, *in vitro* transcribed mRNA encoding for the intron and the IEP can be transformed and then be expressed *in vivo*. Selection for the desired genotype allows for the identification of the correct transformants. (III) A different approach could be to use a plasmid encoding only for the intron, modify and *in vitro* transcribe it. Combining the intronic RNA with recombinantly purified IEP *in vitro* leads to the formation of the active RNP [300]. This RNP consists of the IEP and the intron in its lariat form. Subsequent transformation of the RNP allows for the desired gene modification. In this case, *in vivo* expression of the IEP is not necessary, while also providing the option to modify the IEP, for example with a nuclear localisation tag for usage in eukaryotes [304]. **B)** Scheme for MS2rep dependent amplification of the split EcI5/IEP mRNA *in vivo*. Translation from the $(+)_{rep}$ strand, corresponding to the strand that serves as mRNA for MS2rep, leads to the synthesis of the active MS2rep holo complex. This strand also serves as template for (+) to (-) and (-) to (+) transcription. Translation of the IEP from and splicing of the intron out of the $(-)_{rep}$ strand lead to formation of the active EcI5-RNP. In parallel, the two sequential splicing events lead to the recombination of the 5'- and 3'-end of the $(-)_{rep}$ strand, thereby forming a shortened $(-)_{rep}$. This nascent strand serves as as template for (-) to (+) transcription and vice versa, as well as one of two new mRNA species, encoding for the IEP and MS2rep. Ensuing replication and translation events then effectively increase the copy number of the EcI5-RNP.

Subsequent replication of this shortened $(-)_{rep}$ generates a new mRNA for the replicase. Thus, expression of both MS2rep and IEP would be boosted, the copy number of all four strands would increase through replication and thereby the overall chance of genomic insertion from the EcI5-RNP as well.

Furthermore, combining expression of the EcI5/IEP-RNA from a plasmid in combination with transformation of a RNA replisome, or vice versa with the transformation of the EcI5/IEP-RNP, could also allow the knock-out of essential genes, provided the respective gene can be expressed *in trans* from the replisome. Thereby, expression of this essential gene from the RNA would serve as selection pressure to maintain a functional CDS, while also being useful for potential directed *in vivo* evolution of the gene. Supplementing this approach with bioinformatic methods for the design of overlapping ORFs [305–307] and /or riboswitches [308,309] could further increase the range of possible applications, by increasing the persistence of encoded genes and of the replisome, and conditional expression of these genes, respectively.

4.3 MS2 replicase – Old questions, new answers

4.3.1 MS2rep and Qβrep – Same but different

Already in the 1970s, EF-Tu, EF-Ts and S1 were identified as essential host factors for the replicases of the viruses MS2 and Qβ [66,101,268,310]. In contrast, while Qβ was shown to rely on Hfq (Qβ host factor) *in vivo* [86,311], MS2 was seemingly independent [312]. As it was not possible to isolate active MS2 replicase [117], it was believed that there might be a missing MS2 specific host factor, similar to Hfq.

In this study, the purification of both replicases and reconstitution of the active MS2 complex, highlighted further similarities and differences between MS2 and Qβ. Thus, the MS2 replicase subunit was isolated as an inactive heterotrimeric complex (MS2rep·S1·EF-Ts), but the Qβ replicase subunit was readily obtained as active tetrameric complex (Qβrep·EF-Ts·EF-Tu·S1). However, even though EF-Tu is only weakly bound in the MS2rep complex, activity of the holo complex appeared to be more robust compared to Qβrep. In case of Qβrep, activity was drastically reduced when the N-terminus was tagged, or purification buffers contained EDTA instead of Mg^{2+}. These effects were not observed for MS2rep, indicating that the two replicase complexes are structurally distinct. The two subunits exhibit homology in their core region, the termini however are less conserved, with a sequence identity of 38% (pairwise alignment, BLAST 2 Sequences [313,314]) and a size difference of 41 amino acids (Qβ: 586 aa / MS2: 545 aa). Presumably, the elongated N-terminus of the Qβ replicase subunit plays a crucial role in complex assembly and/or substrate binding, potentially depending on Mg^{2+} as a co-factor. In case of the MS2rep complex, this function might then either not be present, or taken over by a different factor.

This difference might also be reflected in the requirements on putative RNA substrates. The

results presented in this study suggest that the two replication machineries might be at least partially orthogonal to each other, as the Qβrep substrate RQ-135 could not be amplified by MS2rep. Previous findings further support this hypothesis [315,316].

Finally, IF1 was identified as enhancer of MS2rep activity. As far as is known, this was never observed for Qβrep. Therefore, IF1 could potentially be identical with the long missing MS2-specific host factor.

4.3.2 How replication of MS2 could be regulated in vivo

The initiation factors IF1 and IF3 demonstrated substantial influence on MS2rep activity *in vitro*. While IF1 acted as an enhancer, IF3 strongly down regulated replicase activity. For both, RNA binding could be shown. However, these two distinct effects could not be observed *in vivo,* and a potential biological role remained unclear. As already mentioned, the current findings must be taken with a grain of salt, as the method for the *in vivo* detection had several inadequacies. In the following, a recently speculated role [100] of IF1 and IF3 in the *in vivo* life cycle of the MS2 phage will be elaborated.

Part of the activity of IF1, just like for Hfq and S1, stems from oligo binding (OB) folds, a protein structure known for RNA binding [279,317–319]. In case of IF1, this RNA binding enables IF1 to participate in the role of a chaperone in transcription anti-termination [280] and trans-splicing of group I introns [279]. Thereby, IF1 can destabilise secondary structures. For the RNA replication by MS2rep, this presumably results in increased accessibility of recognition/binding sites, a facilitated readthrough, replication initiation and/or dissociation of the product strand. During the transcription of antigenomic from the genomic Qβ RNA, Hfq and S1 take on a similar role. Hereby, the 3'-end binding by the replicase subunit is facilitated by Hfq [86,311], and termination and re-initiation presumably by S1 [320]. A potential overlap of the function of Hfq and IF1 also explains the reported independence of MS2 from Hfq [312]. In mutants lacking a functional *hfq* gene, IF1 could thus potentially compensate for the lost functionality.

The structural resemblance between tRNAs and the UTRs of MS2, could serve a similar purpose, by attracting further RNA binding proteins, or even be directly involved in these effects [125,277,278]. Follow-up studies on this specific question might deepen the understanding of their role in the MS2 life cycle. Furthermore, those results could contribute to elucidate the evolutionary origin and role of tRNA like structures in other RNA species [63,64].

Although both IF1 and IF3 bind to MS2 RNA, they exhibited completely opposite effects on replicase activity. As was reported, IF3 binds to the 3'-UTR of genomic MS2 RNA [281–283]. Here, it was possible to also demonstrate binding of IF3 to antigenomic MS2-like RNA ($[F30\text{-}Bro]_{UTRs(-)}$), with IF3 binding the RNA at multiple sites. However, in contrast to IF1, the binding of the 3'-UTR by IF3 inhibits replication, apparently through a direct competition with the replicase subunit. Interestingly, the expression of MS2 genes *in vitro* was shown to be

independent of IF3 [138]. This suggests that the translation and replication of genomic MS2 RNA are most prominent, when intracellular IF3 concentrations are lowest. Low concentrations of IF3 were reported for *E. coli* cells in the stationary growth phase [321]. It was shown that MS2 virions still form during the stationary phase, albeit this is not followed by cell lysis [322]. In addition, at least one binding site of IF3 in the 3'-UTR (3437 – 3501) [282] either directly overlaps with, or is near binding sites of the coat protein (3400 – 3500) [106] and maturation protein (3540 – 3563), respectively [105]. Capsid formation might therefore be hindered under conditions with high levels of IF3. Finally, IF3 fully inhibited the replicase activity in the *in vitro* assays using the minimal buffer system. However, in the PURExpress and PURE 3.0 systems, which contain IF3 in addition to numerous other proteins, the replicase activity was not as drastically reduced, presumably due to competition with other RNA binding factors. These findings suggest that replication of MS2 RNA is continuous throughout the different phases of bacterial growth, but might be favoured during stationary phase, when resources for translation are limited. This potentially increases persistence of MS2 phage in phases where propagation of progeny virions is impeded and ensure that, once the conditions change, lysis, phage release and new infections take place.

In summary, the three competing key processes – replication, translation, and packaging – are connected to IF3. Thus, levels of IF3 could orchestrate the transition between an open and a closed state (Figure 34). In the open state, IF3 binding down regulates packaging and replication thus allowing for prolonged phases of translation from one individual genomic strand. In the closed state, the absence of IF3 no longer delays packaging and replication of genomic MS2 RNA. Subsequent binding by CP dimers and/or complete packaging protects RNA against RNAse activity and hydrolysis. In addition, higher levels of RNA replication ensure sufficient RNA is available for packaging, but also for protein synthesis once the states switch again.

Unfortunately, both IF1 and IF3 are essential for the survival of *E. coli*. Consequently, a deletion of either of these factors poses a substantial challenge. Follow up studies to verify this hypothesis therefore depend on conditional mutants of these two factors.

4.3.3 MSRP-22 – A new Spiegelman's monster

More than 50 years ago, serial transfer experiments with Qβ replicase, performed in the lab of Sol Spiegelman, demonstrated the generation of short replicating RNAs from genomic Qβ RNA [284,323]. This RNA was then later referred to as Spiegelman's monster. While the initial publication also mentions isolation of, and RNA replication by MS2 replicase [284], the work on MS2rep was abandoned in favour of Qβrep in the time following. This was due to limitations with purification methods at that time [323,324].

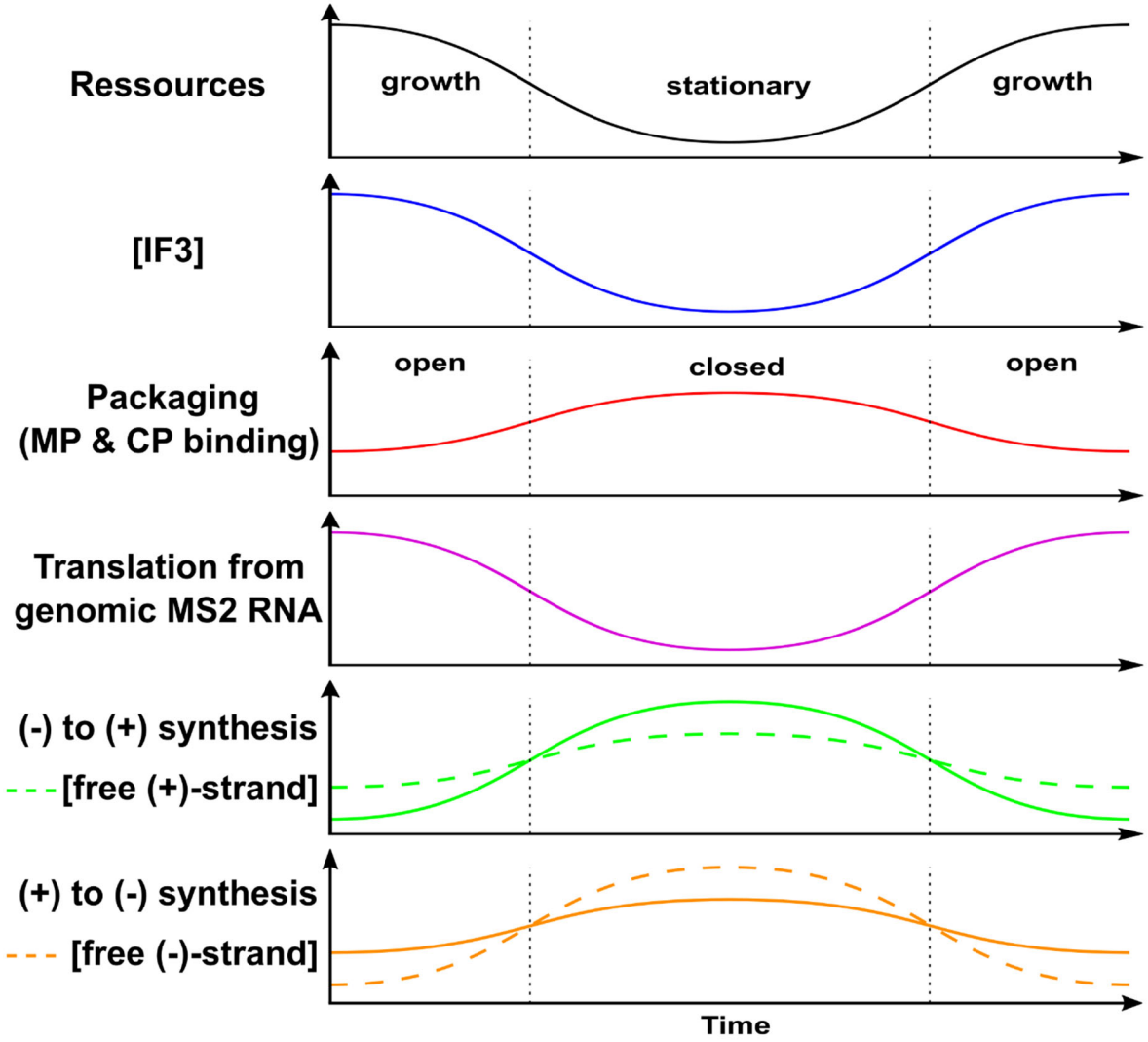


Figure 34 IF3 dependent coordination of MS2 life cycle: Growth rate of *E. coli* cultures correlates with the available resources. Low resources lead to a transition from growth phase to stationary phase. Levels of IF3 and thus levels of general translation of host proteins follow this pattern accordingly. Absence of IF3 shifts the equilibrium from open to closed state, i.e., genomic MS2 RNA can now be more readily packaged into capsids and replicase activity potentially increases. This reduces the competing process of translation, further affected by the decrease of overall translation levels. As packaging depends on synthesis of the capsid components, the decrease of translation counteracts the shift to increased RNA packaging. Thus, the difference in packaging rates presumably is only moderate, which potentially regulates the two now dominating processes – packaging and replication. Furthermore, lower translation levels also result in reduced expression of the *lysis* gene, corresponding to abolished cell lysis [128]. Consequently, concentration of 'free' intracellular (+)-strand, unbound by IF3, ribosomes or packaging factors, increases. However, the increase in 'free' (+)-strand does not fully follow increased levels of (-) to (+) synthesis, due to the packaging of it into capsids, but also more frequent binding of coat protein to the regulator hairpin for replicase expression [124,125,203]. In contrast, increased transcription from (+)-strand leads to a strong increase of free (-)-strand. As the (-)-strand does neither serve as mRNA or substrate for packaging, it can fully serve as template for (+)-strand synthesis. This increase potentially balances packaging of and transcription from (+)-strand. In total, starved bacterial cultures would exhibit higher levels of intracellular MS2 RNA and virions while lysis is not taking place. Due to increased RNA levels and continuous, lysis-independent release of mature virions for background levels of infection, persistence of MS2 within the culture is enabled [322,325,326]. Once the available resources increase again, the lytic phase can be re-established.

Here, the *in vitro* reconstitution of active MS2rep holo complex finally allowed for serial transfer experiments with MS2.

Similar to the Qβ replicase, the reconstituted MS2 replicase was capable to generate short replicative RNA species, even in the absence of a designated template RNA [285,288]. The initial RNA species were only approximately 200 to 300 nt long, however, they quickly diverged into RNAs covering a broader size range. This size increase presumably results from a facilitation of replication initiation. Alternatively, it might be the consequence of RNA recombination, which was also shown for Qβ [286].

Sequencing of this short replicating RNA, named MSRP-22, revealed that it consists only of the native MS2 5'- and 3'-UTRs. This indicated that they comprise the minimally required motifs for RNA replication. Indeed, MSRP-22 demonstrated great potential as a scaffold for replicating RNAs. By incorporating the MS2 *rep* gene, it was even possible to design functional self-replicating RNAs. Furthermore, these replisomes were not only active *in vitro* but also boosted Zeocin resistance in *E. coli* cells *in vivo*, when encoded in a plasmid. In conclusion, they constitute a promising platform for the development of new strategies for *in vitro* and *in vivo* evolution studies, as well as self-replicating RNA genomes.

One interesting aspect of self-replicating RNAs is the design of replication networks (Figure 35). Based on RQ-135 and the Qβ replicase subunit, a second replisome, [*Qβrep*]$_{RQ-135}$, could be generated. Similar Qβ replisomes already exist, albeit based on the Qβ substrate midivariant 1 (MDV-1) [327]. With [*MS2rep*]$_{MSRP-22}$ and [*Qβrep*]$_{RQ-135}$, two presumably orthogonal replication systems would be at hand. Thereby, replication and translation from one RNA/replicase pair would be independent of the other pair (Figure 35A).

Furthermore, both replicators can be coupled to secondary replicators, similar to [F30-Bro]$_{MSRP-22-1.0}$ in the co-replication assays in this study (Figure 29B). In combination with riboswitches, this setup would already allow for design of complex RNA based logic gates [257,308,309]. However, by intertwining the two replicating systems, the complexity can be further increased.

Thus, in the simplest combination, the translation/replication of pair A would depend on pair B (Figure 35B). Here, the translation of the Qβ replicase subunit would depend on prior expression of MS2 replicase subunit and transcription of its respective mRNA. The replication of MS2rep mRNA in turn would depend on the preceding transcription and expression of the Qβrep mRNA. Alternatively, this dependence can also be established in a hybrid system, where one single RNA replisome encodes both replicase, with each only performing one designated half of the replication cycle (Figure 35C).

The potential of these replication networks for the research and design of synthetic gene circuits and future *in vivo* applications is substantial [96,97,328–330]. They also could contribute to provide insights into the early stages of the emergence of life. Thus, *Ichihashi* and colleagues

were already able to demonstrate the emergence of cooperating replication networks *in vitro*, a state similar to the conditions assumed for the origin of life [285,287,331,332].

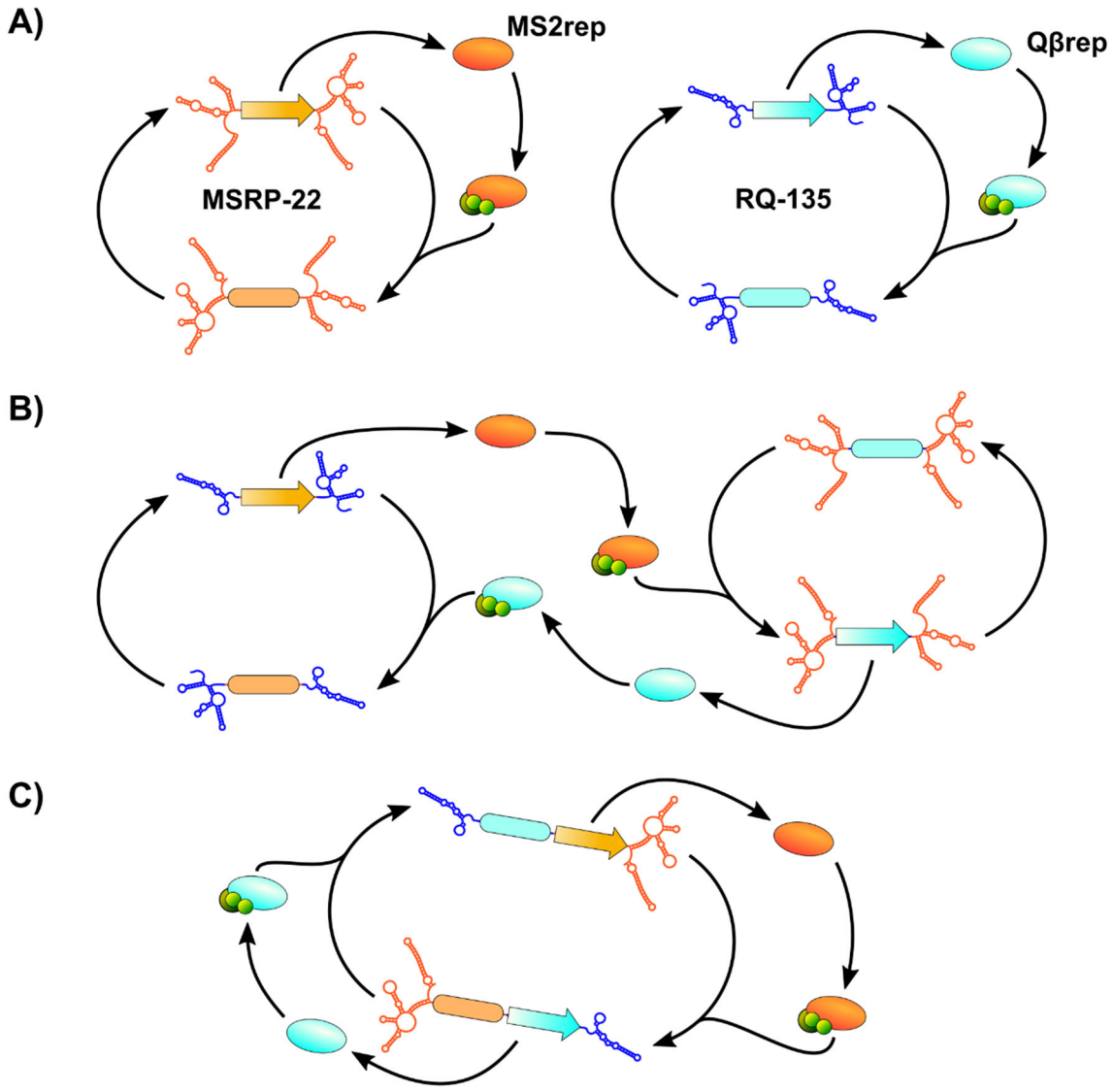


Figure 35 Replication networks: A) Two independent replicator systems, based on MSRP-22 with MS2rep (orange) and RQ-135 with Qβrep (blue). The (+)-strands of each replicator serve as mRNA for the expression of the corresponding replicase subunit and template for RNA replication. The two systems are orthogonal to each other, there is no inter-systemic crosstalk. **B)** Inter-systemic cross-replication of MSRP-22 and RQ-135 based replicators. The (+)-strand of each replicator serves as mRNA for the expression of the non-related replicase subunit. Thus, MSRP-22(+) encodes for Qβrep and RQ-135(+) for MS2rep. The replication of each strand depends on *in trans* cooperation with the respective other strand. **C)** Intra-systemic cross-replication. The replicator is a hybrid of the MSRP-22 and the RQ-135 replicators. Each strand hereby encodes for one of the two distinct replicase subunits and the corresponding 3'-UTR required for the synthesis of its complementary strand. Cooperation between the two replicators occurs *in cis*.

4.4 Conclusion and Future Prospect

At the beginning of this theses stood the idea of harnessing the infection and replication machineries of bacteriophage MS2 to design a self-replicating RNA genome. In consequence, this meant to reprogram a parasitic RNA species into an RNA, that benefits the infected host, analogous to a mutualist. With the discovery of the short replicating RNA MSRP-22, this goal became possible. It did not only exhibit full replication by MS2rep, but also served as a scaffold for diverse genetic information. Furthermore, it was possible to reconstitute minimal infectious units of MS2 *in vitro*. Unfortunately, this did not result in an efficient MP-based system for RNA delivery. But it was possible to establish an alternative approach for this endeavour that was not only highly efficient and reproducible but also low in requirements. In addition, several methods were repurposed for the attempt to reprogramm MS2. While not all where successfully implemented in this thesis to create a self-replicating RNA genome, they exhibited great potential for different problems, as well as for the reprogramming of RNA viruses in general.

However, this study was always part of a bigger project, aiming at the generation of an RNA genome. Hereby, the coding sequences of endogenous host factors were also supposed to be 'transplanted' from the DNA into the RNA genome.

While at the beginning only a single gene would be transplanted, subsequent steps were thought to increase the information exclusively encoded in the RNA genome. Due to the limited stability of RNA, as well as error rate of viral replicases, it was speculated that self-replicating RNAs possess an upper size [61,93,94]. The currently largest known RNA virus is *planarian secretory cell nidovirus* (PSCNV), with a single genomic RNA strand of 41.1 kb [93]. Including all biological pathways involved in DNA synthesis and maintenance, the proposed size of a minimal genome is around 113 kb [99], thus still almost three time bigger than the PSCNV genome.

Nonetheless, a way how RNA viruses in nature increase their genome size, is by segmentation. Here, the genetic information is spread among several RNA chromosomes, which are replicated by the viral replicase [333,334]. This is the case for *Orthomyxoviridae*, like Influenza viruses causing the common flu, *Arenaviridae*, like the hemorrhagic fever causing *Lassa virus*, or the plant pathogenic *Bromoviridae* [61]. In this study, it was already demonstrated that the MS2 replicase complex is capable of co-replication of at least a second 'RNA chromosome' (Figure 29B). Thus, it fulfils an important prerequisite for the design of large, segmented RNA genomes. Combination of these putative segments with packaging RNAs (pRNA) [335–337] could provide a way to ensure stoichiometry and co-transmission of the individual segments during cell division (Figure 36). Alternatively, or even as addition, partitioning systems as they are known for DNA plasmids might be adapted to contribute to the maintenance and inheritance of these RNAs [338,339]. In the long-term, it could even serve as

a steppingstone for the design of DNA-free minimal cells, by pushing the size limitations of RNA genomes.

Finally, although this work might have answered some questions regarding the MS2 life cycle and demonstrated diverse aspects of RNA replication, it was only a step towards novel RNA genomes, with high prospect for novel insights and future applications.

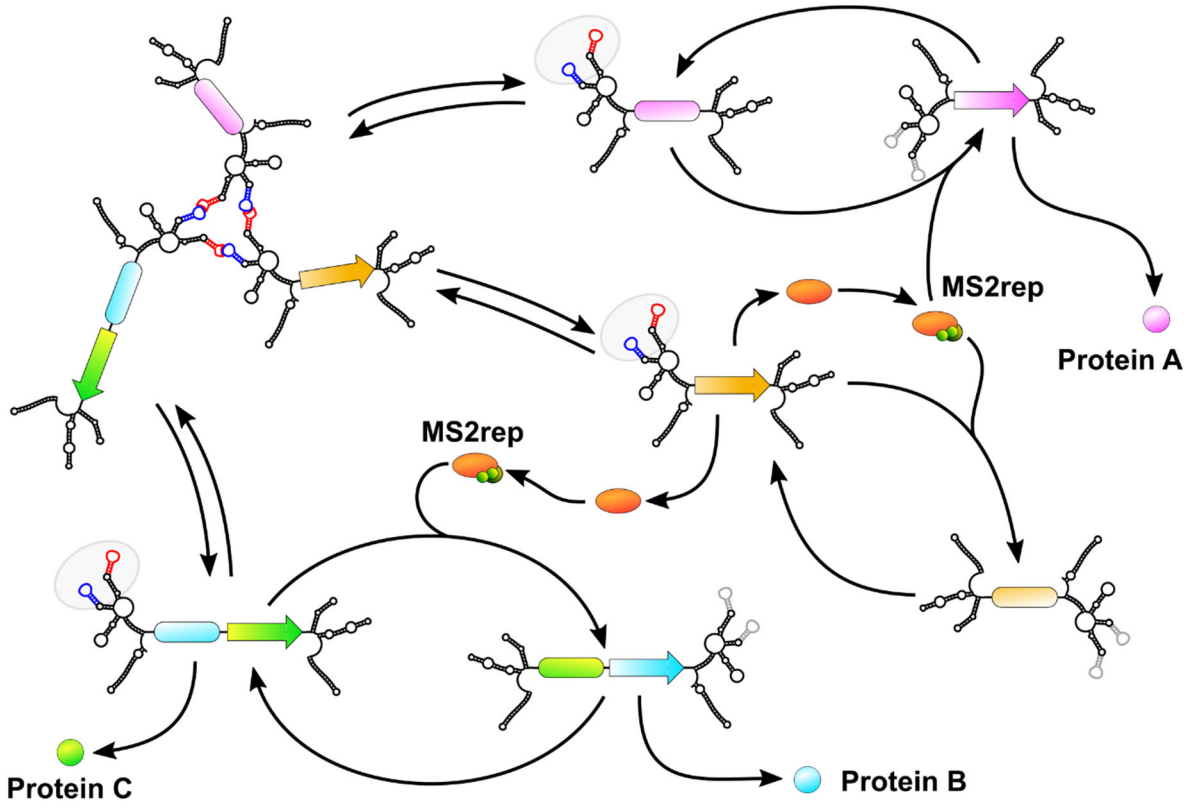


Figure 36 Example for a replication network with the potential to form higher ordered structures: The replication network consists of individual replicators, each encoding for a distinct set of genes of interest, and for the MS2 replicase subunit. Expression of the *rep* gene is followed by the formation of the active replicase holo complex (MS2rep). The replicase can then replicate the individual strands through their corresponding complementary strand. In addition, pRNA derived structures allow the individual strands to assemble into a higher order structure of well-defined composition. Thus, transfer to daughter cells of the distinct strands in fixed stoichiometry is possible.

5. Acknowledgments

This is a thank you, to all the great people that made this possible.

Hannes Mutschler, **thank you very much** for taking me in as a PhD student, for letting me work on this amazing project, for allowing me to do my sometimes-crazy side projects and for your great mentorship and guidance. **Thank you very, very much**.

Klaus Förstemann, **thank you very much** for your supervision and guidance, both throughout my time as PhD student, as well as my time as a Bachelor and Master student. **Thank you very much** for being part of my TAC and providing help and insights of immeasurable value.

Petra Schwille, **thank you very much** for covering for Hannes as official Doktormutter, making it possible that I could even start this journey.

Dany Nedialkova, **thank you very much**. Your tips and recommendations during my TAC meetings helped me focus and prevented me getting lost in subprojects.

Laura Weise and Kai Libicher, **thank you very much** for your friendship and the great talks, science or not. This whole thing would not have come to be without all your pre-work and contributions.

Victoria, Renate, and Lena, **thank you very much** for all your assistance and help in the lab and beyond. Your work made this

Alexander Floroni, **thank you very much** for volunteering to be supervised by me, directly contributing to this work and all your curiosity and cheerfulness.

Deni, Elia, Emilie, Jacopo, and Kris, **thank you very, very, very much** for the time we shared together, might it have been in the lab and offices, talking science, or amazing game nights and pub crawls. You guys might have made my productivity suffer at times, but only for the best.

My parents, my sister and everyone else I am lucky enough to consider friends and family, **thank you very much**. You shaped the path that let me here.

Eleni, **thank you very much**, for being here with me and for all the good your brought to my life, the things that can be put in words and those that can’t. **Σ' αγαπώ πολύ**.

And you, **thank you very much** for taking your time and reading this work up to this point. I hope you found it helpful and interesting.

6. References

1. Fitch, W. T. Glossogeny and phylogeny: cultural evolution meets genetic evolution. *Trends in Genetics* **24**, 373–374 (2008).

2. Gingeras, T. R. Origin of phenotypes: Genes and transcripts. *Genome Research* **17**, 682–690 (2007).

3. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).

4. Jaafar, Z. A. & Kieft, J. S. Viral RNA structure-based strategies to manipulate translation. *Nature Reviews Microbiology* **17**, 110–123 (2018).

5. Cooke, M. S., Evans, M. D., Dizdaroglu, M. & Lunec, J. Oxidative DNA damage: mechanisms, mutation, and disease. *The FASEB Journal* **17**, 1195–1214 (2003).

6. Loh, E., Salk, J. J. & Loeb, L. A. Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc Natl Acad Sci U S A* **107**, 1154–1159 (2010).

7. Robertson, D. S. Feedback theory and Darwinian evolution. *Journal of Theoretical Biology* **152**, 469–484 (1991).

8. Petroni, F. & Serva, M. Language distance and tree reconstruction. *Journal of Statistical Mechanics: Theory and Experiment* **2008**, P08012 (2008).

9. Gąsiorowski, P. The Tree of Language: A Cladistic Look at the Genetic Classification of Languages. *DiG* **1999**, 57 (1999).

10. Moreira, D. & Philippe, H. Molecular phylogeny: pitfalls and progress. *Int Microbiol* **3**, 9–16 (2000).

11. Nasir, A. & Caetano-Anollés, G. A phylogenomic data-driven exploration of viral origins and evolution. *Science Advances* **1**, (2015).

12. Harbert, W. *The germanic languages*. *The Germanic Languages* (Cambridge University Press, 2006). doi:10.1017/CBO9780511755071.

13. Hartmann, F. & Riegger, C. The Burgundian language and its phylogeny: A cladistical investigation. *NOWELE North-Western European Language Evolution* **75**, 42–80 (2022).

14. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).

15. Woese, T. C. & Tang, S. The Origin(s) of Cell(s): Pre-Darwinian Evolution from FUCAs to LUCA. *Journal of Molecular Evolution 2021 89:7* **89**, 427–447 (2021).

16. Mushegian, A. Gene content of LUCA, the last universal common ancestor. *Front Biosci* **13**, 4657–4666 (2008).

17. Fine, P. V. A. Ecological and Evolutionary Drivers of Geographic Variation in Species Diversity. *Annual Review of Ecology, Evolution and Systematics* **46**, 369–392 (2015).

18. Prüfer, K. *et al.* A high-coverage Neandertal genome from Vindija Cave in Croatia. *Science (1979)* **358**, 658 (2017).

19. Browning, S. R., Browning, B. L., Zhou, Y., Tucci, S. & Akey, J. M. Analysis of Human Sequence Data Reveals Two Pulses of Archaic Denisovan Admixture. *Cell* **173**, 53-61.e9 (2018).

20. Kurland, C. G., Canback, B. & Berg, O. G. Horizontal gene transfer: A critical view. *Proc Natl Acad Sci U S A* **100**, 9658–9662 (2003).

21. Brown, J. R. Ancient horizontal gene transfer. *Nature Reviews Genetics* **4**, 121–132 (2003).

22. Morshed, S. English Calques in Bangla. *Journal of Language Learning and Research* **2**, 14–23 (2019).

23. Mattiello, E. The Pervasiveness of Slang in Standard and Non-Standard English. *Mots Palabras Words* **6**, (2005).

24. Villarreal, L. P. Evolution of Viruses. *Encyclopedia of Virology* 174 (2008) doi:10.1016/B978-012374410-4.00706-8.

25. Dahm, R. Friedrich Miescher and the discovery of DNA. *Developmental Biology* **278**, 274–288 (2005).

26. Hiyoshi, A., Miyahara, K., Kato, C. & Ohshima, Y. Does a DNA-less cellular organism exist on Earth? *Genes Cells* **16**, 1146–1158 (2011).

27. Robertson, M. P. & Joyce, G. F. The Origins of the RNA World. *Cold Spring Harbor Perspectives in Biology* **4**, a003608 (2012).

28. Joyce, G. F. & Szostak, J. W. Protocells and RNA Self-Replication. *Cold Spring Harbor Perspectives in Biology* **10**, a034801 (2018).

29. Bernhardt, H. S. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others). *Biology Direct* **7**, 23 (2012).

30. Caetano-Anollés, G. & Seufferheld, M. J. The Coevolutionary Roots of Biochemistry and Cellular Organization Challenge the RNA World Paradigm. *Microbial Physiology* **23**, 152–177 (2013).

31. Pearce, B. K. D., Pudritz, R. E., Semenov, D. A. & Henning, T. K. Origin of the RNA world: The fate of nucleobases in warm little ponds. *Proc Natl Acad Sci U S A* **114**, 11327–11332 (2017).

32. Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A. & Holliger, P. Ice as a protocellular medium for RNA replication. *Nat Commun* **1**, (2010).

33. Vlassov, A. v., Kazakov, S. A., Johnston, B. H. & Landweber, L. F. The RNA world on ice: a new scenario for the emergence of RNA information. *J Mol Evol* **61**, 264–273 (2005).

34. Bernhardt, H. S. & Tate, W. P. Primordial soup or vinaigrette: Did the RNA world evolve at acidic pH? *Biology Direct* **7**, 1–12 (2012).

35. Jayasena, V. K. & Gold, L. In vitro selection of self-cleaving RNAs with a low pH optimum. *Proc Natl Acad Sci U S A* **94**, 10612–10617 (1997).

36. Xu, J. *et al.* Selective prebiotic formation of RNA pyrimidine and DNA purine nucleosides. *Nature* **582**, 60–66 (2020).

37. le Vay, K. & Mutschler, H. How DNA and RNA subunits might have formed to make the first genetic alphabet. *Nature* **582**, 33–34 (2020).

38. Gull, M. Prebiotic Phosphorylation Reactions on the Early Earth. *Challenges* **5**, 193–212 (2014).

39. Fialho, D. M., Roche, T. P. & Hud, N. v. Prebiotic Syntheses of Noncanonical Nucleosides and Nucleotides. *Chemical Reviews* **120**, 4806–4830 (2020).

40. Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science (1979)* **261**, 1411–1418 (1993).

41. Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science (1979)* **292**, 1319–1325 (2001).

42. Ruiz-Mirazo, K., Moreno, A., Ruiz-Mirazo, K. & Moreno, · A. Autonomy in evolution: from minimal to complex life. *Synthese* **185**, 21–52 (2011).

43. Ruiz-Mirazo, K. & Moreno, A. Basic Autonomy as a Fundamental Step in the Synthesis of Life. *Artificial Life* **10**, 235–259 (2004).

44. Fusz, S., Eisenführ, A., Srivatsan, S. G., Heckel, A. & Famulok, M. A ribozyme for the aldol reaction. *Chem Biol* **12**, 941–950 (2005).

45. Hiller, D. A. & Strobel, S. A. The chemical versatility of RNA. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**, 2929 (2011).

46. Turk, R. M., Chumachenko, N. v. & Yarus, M. Multiple translational products from a five-nucleotide ribozyme. *Proc Natl Acad Sci U S A* **107**, 4585–4589 (2010).

47. Joyce, G. F. The antiquity of RNA-based evolution. *Nature* **418**, 214–221 (2002).

48. Cochrane, J. C. & Strobel, S. A. Riboswitch effectors as protein enzyme cofactors. *RNA* **14**, 993 (2008).

49. Steitz, T. A. & Moore, P. B. RNA, the first macromolecular catalyst: the ribosome is a ribozyme. *Trends in Biochemical Sciences* **28**, 411–418 (2003).

50. Wolf, Y. I. & Koonin, E. v. On the origin of the translation system and the genetic code in the RNA world by means of natural selection, exaptation, and subfunctionalization. *Biology Direct* **2**, 1–25 (2007).

51. Harish, A. & Caetano-Anollés, G. Ribosomal History Reveals Origins of Modern Protein Synthesis. *PLOS ONE* **7**, e32776 (2012).

52. Bernhardt, H. S. & Tate, W. P. The transition from noncoded to coded protein synthesis: Did coding mRNAs arise from stability-enhancing binding partners to tRNA? *Biology Direct* **5**, 1–18 (2010).

53. de Farias, S. T., Rêgo, T. G. & José, M. v. tRNA Core Hypothesis for the Transition from the RNA World to the Ribonucleoprotein World. *Life* **6**, 15 (2016).

54. Frankel, E. A., Bevilacqua, P. C. & Keating, C. D. Polyamine/Nucleotide Coacervates Provide Strong Compartmentalization of Mg2+, Nucleotides, and RNA. *Langmuir* **32**, 2041–2049 (2016).

55. le Vay, K., Song, E. Y., Ghosh, B., Tang, T. Y. D. & Mutschler, H. Enhanced Ribozyme-Catalyzed Recombination and Oligonucleotide Assembly in Peptide-RNA Condensates. *Angewandte Chemie International Edition* **60**, 26096–26104 (2021).

56. Iranzo, J., Puigbo, P., Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. v. Inevitability of Genetic Parasites. *Genome Biology and Evolution* **8**, 2856–2869 (2016).

57. de Farias, S. T., dos Santos, A. P., Rêgo, T. G. & José, M. v. Origin and evolution of RNA-dependent RNA polymerase. *Frontiers in Genetics* **8**, 125 (2017).

58. Černý, J., Bolfíková, B. Č., Valdés, J. J., Grubhoffer, L. & Růžek, D. Evolution of Tertiary Structure of Viral RNA Dependent Polymerases. *PLOS ONE* **9**, e96070 (2014).

59. de Farias, S. T., Rêgo, T. G. & José, M. v. A proposal of the proteome before the last universal common ancestor (LUCA). *International Journal of Astrobiology* **15**, 27–31 (2016).

60. Schopf, J. W. The First Billion Years: When Did Life Emerge? *Elements* **2**, 229–233 (2006).

61. Reanney, D. C. The Evolution of RNA Viruses. *Annual Review of Microbiology* **36**, 47–73 (1982).

62. Belshaw, R., Pybus, O. G. & Rambaut, A. The evolution of genome compression and genomic novelty in RNA viruses. *Genome Research* **17**, 1496 (2007).

63. Ariza-Mateos, A. & Gómez, J. Viral tRNA mimicry from a biocommunicative perspective. *Frontiers in Microbiology* **8**, 2395 (2017).

64. Wu, S., Li, X. & Wang, G. tRNA-like structures and their functions. *FEBS Journal* (2021) doi:10.1111/FEBS.16070.

65. Louten, J. Virus Structure and Classification. *Essential Human Virology* 19 (2016) doi:10.1016/B978-0-12-800947-5.00002-8.

66. Fiers, W. *et al.* Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260**, 500–507 (1976).

67. Kashiwagi, A. & Yomo, T. Ongoing Phenotypic and Genomic Changes in Experimental Coevolution of RNA Bacteriophage Qβ and Escherichia coli. *PLoS Genetics* **7**, 1002188 (2011).

68. Naqvi, A. A. T. *et al.* Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: Structural genomics approach. *Biochimica et Biophysica Acta. Molecular Basis of Disease* **1866**, 165878 (2020).

69. Cheung, T. K. W. & Poon, L. L. M. Biology of Influenza A Virus. *Ann N Y Acad Sci* **1102**, 1–25 (2007).

70. Jun, S. R. *et al.* Ebolavirus comparative genomics. *FEMS Microbiology Reviews* **39**, 764–778 (2015).

71. Wunner, W. H. & Conzelmann, K.-K. Rabies virus. *Rabies* 43–81 (2020) doi:10.1016/B978-0-12-818705-0.00002-9.

72. Desselberger, U. Rotaviruses. *Virus Research* **190**, 75–96 (2014).

73. Maan, S. *et al.* Full-Genome Sequencing as a Basis for Molecular Epidemiology Studies of Bluetongue Virus in India. *PLOS ONE* **10**, e0131257 (2015).

74. Liao, D. *et al.* Complete Genome Sequence of a Bluetongue Virus Serotype 15 Strain Isolated from China in 1996. *Genome Announcements* **6**, 557–575 (2018).

75. Domingo, E. & Holland, J. J. RNA Virus Mutations and Fitness For Survival. *Annual Review of Microbiology* **51**, 151–178 (2003).

76. Chen, N. *et al.* RNA sensors of the innate immune system and their detection of pathogens. *IUBMB Life* **69**, 297–304 (2017).

77. Wu, F. *et al.* A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).

78. Hoang, H. D., Neault, S., Pelin, A. & Alain, T. Emerging translation strategies during virus–host interaction. *Wiley Interdisciplinary Reviews: RNA* **12**, e1619 (2021).

79. Navarro, B., Flores, R. & di Serio, F. Advances in Viroid-Host Interactions. *Annual Review of Virology* **8**, 305–325 (2021).

80. Raj Adkar-Purushothama, C., Perreault, J.-P., Group, R. & Arn, G. Current overview on viroid–host interactions. *Wiley Interdisciplinary Reviews: RNA* **11**, e1570 (2020).

81. Pallas, V., García-Luque, I., Domingo, E. & Flores, R. Sequence variability in avocado sunblotch viroid (ASBV). *Nucleic Acids Research* **16**, 9864–9864 (1988).

82. Brichot, C. Hepatitis C virus. *Digestive Diseases and Sciences* **41**, 6S-21S (1996).

83. Palukaitis, P. Satellite RNAs and satellite viruses. *Molecular Plant-Microbe Interactions* **29**, 181–186 (2016).

84. Hu, C. C., Hsu, Y. H. & Lin, N. S. Satellite RNAs and Satellite Viruses of Plants. *Viruses* **1**, 1325–1350 (2009).

85. Mizuuchi, R., Ichihashi, N., Usui, K., Kazuta, Y. & Yomo, T. Adaptive evolution of an artificial RNA genome to a reduced ribosome environment. *ACS Synthetic Biology* **4**, 292–298 (2015).

86. Schuppli, D., Georgijevic, J. & Weber, H. Synergism of mutations in bacteriophage Qbeta RNA affecting host factor dependence of Qbeta replicase. *J Mol Biol* **295**, 149–154 (2000).

87. Rumnieks, J. & Tars, K. Diversity of pili-specific bacteriophages: genome sequence of IncM plasmid-dependent RNA phage M. *BMC Microbiol* **12**, 1–8 (2012).

88. Simmonds, P. *et al.* ICTV virus taxonomy profile: Flaviviridae. *Journal of General Virology* **98**, 2–3 (2017).

89. van den Elsen, K., Quek, J. P. & Luo, D. Molecular Insights into the Flavivirus Replication Complex. *Viruses* **13**, 956 (2021).

90. Klema, V. J., Padmanabhan, R. & Choi, K. H. Flaviviral Replication Complex: Coordination between RNA Synthesis and 5'-RNA Capping. *Viruses* **7**, 4640–4656 (2015).

91. Osawa, S., Jukes, T. H., Watanabe, K. & Muto', A. Recent evidence for evolution of the genetic code. *Microbiological Reviews* **56**, 229–264 (1992).

92. Lustig, F. *et al.* Codon discrimination and anticodon structural context. *Proc Natl Acad Sci U S A* **86**, 6873–6877 (1989).

93. Saberi, A., Gulyaeva, A. A., Brubacher, J. L., Newmark, P. A. & Gorbalenya, A. E. A planarian nidovirus expands the limits of RNA genome size. *PLOS Pathogens* **14**, e1007314 (2018).

94. Lai, M. M. C. The making of infectious viral RNA: No size limit in sight. *Proc Natl Acad Sci U S A* **97**, 5025 (2000).

95. Dolgin, E. The tangled history of mRNA vaccines. *Nature* **597**, 318–324 (2021).

96. Bloom, K., van den Berg, F. & Arbuthnot, P. Self-amplifying RNA vaccines for infectious diseases. *Gene Therapy* **28**, 117–129 (2020).

97. Tews, B. A. & Meyers, G. Self-Replicating RNA. *RNA Vaccines* **1499**, 15 (2017).

98. Yoshioka, N. *et al.* Efficient Generation of Human iPSCs by a Synthetic Self-Replicative RNA. *Cell Stem Cell* **13**, 246–254 (2013).

99. Forster, A. C. & Church, G. M. Towards synthesis of a minimal cell. *Molecular Systems Biology* **2**, 45 (2006).

100. Wagner, A., Weise, L. I. & Mutschler, H. In vitro characterisation of the MS2 RNA polymerase complex reveals host factors that modulate emesviral replicase activity. *Communications Biology* **5**, 1–10 (2022).

101. Blumenthal, T., Landers, T. A. & Weber, K. Bacteriophage Q replicase contains the protein biosynthesis elongation factors EF Tu and EF Ts. *Proc Natl Acad Sci U S A* **69**, 1313–1317 (1972).

102. Takeshita, D. & Tomita, K. Assembly of Qβ viral RNA polymerase with host translational elongation factors EF-Tu and -Ts. *Proc Natl Acad Sci U S A* **107**, 15733–15738 (2010).

103. Callanan, J. *et al.* Leviviricetes: expanding and restructuring the taxonomy of bacteria-infecting single-stranded RNA viruses. *Microbial Genomics* **7**, 686 (2021).

104. Atkins, J. F., Steitz, J. A., Anderson, C. W. & Model, P. Binding of mammalian ribosomes to ms2 phage rna reveals an overlapping gene encoding a lysis function. *Cell* **18**, 247–256 (1979).

105. Dai, X. *et al.* In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **541**, 112–116 (2017).

106. Rolfsson, Ó. *et al.* Direct Evidence for Packaging Signal-Mediated Assembly of Bacteriophage MS2. *Journal of Molecular Biology* **428**, 431–448 (2016).

107. Meng, R. *et al.* Structural basis for the adsorption of a single-stranded RNA bacteriophage. *Nature Communications* **10**, 1–8 (2019).

108. Lawley, T. D., Klimke, W. A., Gubbins, M. J. & Frost, L. S. F factor conjugation is a true type IV secretion system. *FEMS Microbiology Letters* **224**, 1–15 (2003).

109. Christie, P. J., Whitaker, N. & González-Rivera, C. Mechanism and structure of the bacterial type IV secretion systems. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1843**, 1578–1591 (2014).

110. Hu, B., Khara, P. & Christie, P. J. Structural bases for F plasmid conjugation and F pilus biogenesis in Escherichia coli. *Proc Natl Acad Sci U S A* **116**, 14222–14227 (2019).

111. O'Callaghan, R., Bradley, R. & Paranchych, W. The effect of M13 phage infection upon the F pili of E. coli. *Virology* **54**, 220–229 (1973).

112. Clarke, M., Maddera, L., Harris, R. L. & Silverman, P. M. F-pili dynamics by live-cell imaging. *Proc Natl Acad Sci U S A* **105**, 17978–17981 (2008).

113. Harb, L. *et al.* ssRNA phage penetration triggers detachment of the F-pilus. *Proc Natl Acad Sci U S A* **117**, 25751–25758 (2020).

114. Zinder, N. D. & Cooper, S. Host-dependent mutants of the bacteriophage f2 I. Isolation and preliminary classification. *Virology* **23**, 152–158 (1964).

115. Sugiyama, T. & Nakada, D. Translational control of bacteriophage MS2 RNA cistrons by MS2 coat protein: Polyacrylamide gel electrophoretic analysis of proteins synthesized in vitro. *Journal of Molecular Biology* **31**, 431–440 (1968).

116. Cooper, S. & Zinder, N. D. The growth of an RNA bacteriophage: The role of protein synthesis. *Virology* **20**, 605–612 (1963).

117. Fedoroff, N. v. & Zinder, N. D. Factor Requirement of the Bacteriophage f2 Replicase. *Nature New Biology* **241**, 105–108 (1973).

118. Dykeman, E. C., Stockley, P. G. & Twarock, R. Packaging signals in two single-stranded RNA viruses imply a conserved assembly mechanism and geometry of the packaged genome. *Journal of Molecular Biology* **425**, 3235–3249 (2013).

119. Chamakura, K. R., Tran, J. S. & Young, R. MS2 Lysis of Escherichia coli Depends on Host Chaperone DnaJ. *Journal of Bacteriology* **199**, (2017).

120. Chamakura, K. R., Edwards, G. B. & Young, R. Mutational analysis of the MS2 lysis protein L. *Microbiology (N Y)* **163**, 961–969 (2017).

121. Bernhardt, T. G., Wang, I. N., Struck, D. K. & Young, R. Breaking free: "Protein antibiotics" and phage lysis. *Research in Microbiology* **153**, 493–501 (2002).

122. de Crouy-Chanel, A., Kohiyama, M. & Richarme, G. A novel function of Escherichia coli chaperone DnaJ Protein-disulfide isomerase. *J Biol Chem* **270**, 22669–22672 (1995).

123. Yamamoto, K. & Yoshikura, H. Relation between genomic and capsid structures in RNA viruses. *Nucleic Acids Research* **14**, 389–396 (1986).

124. Peabody, D. S. Role of the coat Protein-RNA interaction in the life cycle of bacteriophage MS2. *Molecular and General Genetics MGG* **254**, 358–364 (1997).

125. Berkhout, B. & van Duin, J. Mechanism of translational coupling between coat protein and replicase genes of RNA bacteriophage MS2. *Nucleic Acids Research* **13**, 6955–6967 (1985).

126. Groeneveld, H., Thimon, K. & van Duin, J. Translational control of maturation-protein synthesis in phage MS2: a role for the kinetics of RNA folding? *RNA* **1**, 79 (1995).

127. de Smit, M. H. & van Duin, J. Translational initiation at the coat-protein gene of phage MS2: native upstream RNA relieves inhibition by local secondary structure. *Molecular Microbiology* **9**, 1079–1088 (1993).

128. Schmidt, B. F., Berkhout, B., Overbeek, G. P., van Strien, A. & van Duin, J. Determination of the RNA secondary structure that regulates lysis gene expression in bacteriophage MS2. *Journal of Molecular Biology* **195**, 505–516 (1987).

129. Kastelein, R. A., Remaut, E., Fiers, W. & van Duin, J. Lysis gene expression of RNA phage MS2 depends on a frameshift during translation of the overlapping coat protein gene. *Nature* **295**, 35–41 (1982).

130. Rolfsson, O., Toropova, K., Ranson, N. A. & Stockley, P. G. Mutually-induced Conformational Switching of RNA and Coat Protein Underpins Efficient Assembly of a Viral Capsid. *Journal of Molecular Biology* **401**, 309–322 (2010).

131. Borodavka, A., Tuma, R. & Stockley, P. G. Evidence that viral RNAs have evolved for efficient, two-stage packaging. *PNAS* **109**, (2012).

132. Jou, W. M., Haegeman, G., Ysebaert, M. & Fiers, W. Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein. *Nature* **237**, 82–88 (1972).

133. Sanger, F. *et al.* Nucleotide sequence of bacteriophage φX174 DNA. *Nature* **265**, 687–695 (1977).

134. Nurk, S. *et al.* The complete sequence of a human genome. *Science (1979)* **376**, 44–53 (2022).

135. Strauss, J. H. & Sinsheimer, R. L. Purification and properties of bacteriophage MS2 and of its ribonucleic acid. *Journal of Molecular Biology* **7**, 43–54 (1963).

136. Kastelein, R. A., Berkhout, B., Overbeek, G. P. & van Duin, J. Effect of the sequences upstream from the ribosome-binding site on the yield of protein from the cloned gene for phage MS2 coat protein. *Gene* **23**, 245–254 (1983).

137. Poot, R. A., Tsareva, N. v., Boni, I. v. & van Duin, J. RNA folding kinetics regulates translation of phage MS2 maturation gene. *Proc Natl Acad Sci U S A* **94**, 10110–10115 (1997).

138. Zipori, P., Bosch, L. & van Duin, J. Translation of MS2 RNA in vitro in the absence of initiation factor IF-3. *Eur J Biochem* **92**, 235–241 (1978).

139. Wang, S. *et al.* Preparation and evaluation of MS2 bacteriophage-like particles packaging hepatitis E virus RNA. *FEMS Microbiol Lett* **363**, (2016).

140. Kenyon, J. C., Prestwood, L. J. & Lever, A. M. L. A novel combined RNA-protein interaction analysis distinguishes HIV-1 Gag protein binding sites from structural change in the viral RNA leader. *Sci Rep* **5**, (2015).

141. Dawson, D. J., Paish, A., Staffell, L. M., Seymour, I. J. & Appleton, H. Survival of viruses on fresh produce, using MS2 as a surrogate for norovirus. *Journal of Applied Microbiology* **98**, 203–209 (2005).

142. Mikel, P. *et al.* Preparation of MS2 phage-like particles and their use as potential process control viruses for detection and quantification of enteric RNA viruses in different matrices. *Frontiers in Microbiology* **7**, 1911 (2016).

143. Miranda, J. A. & Steward, G. F. Variables influencing the efficiency and interpretation of reverse transcription quantitative PCR (RT-qPCR): An empirical study using Bacteriophage MS2. *Journal of Virological Methods* **241**, 1–10 (2017).

144. Bertrand, E. *et al.* Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell* **2**, 437–445 (1998).

145. Zhan, S. *et al.* Armored long RNA controls or standards for branched DNA assay for detection of human immunodeficiency virus type 1. *Journal of Clinical Microbiology* **47**, 2571–2576 (2009).

146. Pasloske, B. L., Walkerpeach, C. R., Obermoeller, R. D., Winkler, M. & Dubois, D. B. Armored RNA technology for production of ribonuclease-resistant viral RNA controls and standards. *Journal of Clinical Microbiology* **36**, 3590–3594 (1998).

147. Galaway, F. A. & Stockley, P. G. MS2 viruslike particles: A robust, semisynthetic targeted drug delivery platform. *Molecular Pharmaceutics* **10**, 59–68 (2013).

148. Kovacs, E. W. *et al.* Dual-Surface-Modified Bacteriophage MS2 as an Ideal Scaffold for a Viral Capsid-Based Drug Delivery System. *Bioconjugate Chemistry* **18**, 1140–1147 (2007).

149. Fu, Y. & Li, J. A novel delivery platform based on Bacteriophage MS2 virus-like particles. *Virus Research* **211**, 9–16 (2016).

150. Peabody, D. S., Peabody, J., Bradfute, S. B. & Chackerian, B. RNA Phage VLP-Based Vaccine Platforms. *Pharmaceuticals* **14**, 764 (2021).

151. Li, J. *et al.* Messenger RNA vaccine based on recombinant MS2 virus-like particles against prostate cancer. *International Journal of Cancer* **134**, 1683–1694 (2014).

152. Frietze, K. M., Peabody, D. S. & Chackerian, B. Engineering virus-like particles as vaccine platforms. *Current Opinion in Virology* **18**, 44–49 (2016).

153. Villarreal, L. P. Are Viruses Alive? *Sci Am* **291**, 100–105 (2004).

154. Moreira, D. & López-García, P. Ten reasons to exclude viruses from the tree of life. *Nature Reviews Microbiology* **7**, 306–311 (2009).

155. Regenmortel, M. H. v. *et al. Virus taxonomy: classification and nomenclature of viruses. Seventh report of the International Committee on Taxonomy of Viruses.* (Academic Press, 1999).

156. Raoult, D. *et al.* The 1.2-megabase genome sequence of Mimivirus. *Science (1979)* **306**, 1344–1350 (2004).

157. Pérez-Brocal, V. *et al.* A small microbial genome: The end of a long symbiotic relationship? *Science (1979)* **314**, 312–313 (2006).

158. Nakabachi, A. *et al.* The 160-kilobase genome of the bacterial endosymbiont Carsonella. *Science (1979)* **314**, 267 (2006).

159. Koonin, E. v. & Starokadomskyy, P. Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **59**, 125–134 (2016).

160. Edwards, R. A. & Rohwer, F. Viral metagenomics. *Nature Reviews Microbiology* **3**, 504–510 (2005).

161. Lawrence, C. M. *et al.* Structural and Functional Studies of Archaeal Viruses. *Journal of Biological Chemistry* **284**, 12599–12603 (2009).

162. Nee, S. & Smith, J. M. The evolutionary biology of molecular parasites. *Parasitology* **100**, S5–S18 (1990).

163. Sibley, L. D. Intracellular Parasite Invasion Strategies. *Science (1979)* **304**, 248–253 (2004).

164. Häring, M. *et al.* Independent virus development outside a host. *Nature* **436**, 1101–1102 (2005).

165. Toribio, R. & Ventoso, I. Inhibition of host translation by virus infection in vivo. *Proc Natl Acad Sci U S A* **107**, 9837–9842 (2010).

166. Bird, S. W. & Kirkegaard, K. Escape of non-enveloped virus from intact cells. *Virology* **479–480**, 444–449 (2015).

167. Ampomah, P. B. & Lim, L. H. K. Influenza A virus-induced apoptosis and virus propagation. *Apoptosis* **25**, 1–11 (2020).

168. Getz, W. M. Biomass transformation webs provide a unified approach to consumer-resource modelling. *Ecology Letters* **14**, 113–124 (2011).

169. Savoy, S. K. A. & Boudreau, J. E. The Evolutionary Arms Race between Virus and NK Cells: Diversity Enables Population-Level Virus Control. *Viruses* **11**, 959 (2019).

170. Iranzo, J., Lobkovsky, A. E., Wolf, Y. I. & Koonin, E. v. Virus-host arms race at the joint origin of multicellularity and programmed cell death. *Cell Cycle* **13**, 3083–3088 (2014).

171. Coffin, J. M. Virions at the Gates: Receptors and the Host–Virus Arms Race. *PLOS Biology* **11**, e1001574 (2013).

172. Martin, B. D. & Schwab, E. Current Usage of Symbiosis and Associated Terminology. *International Journal of Biology* **5**, (2013).

173. Obeng, N., Pratama, A. A. & Elsas, J. D. van. The Significance of Mutualistic Phages for Bacterial Ecology and Evolution. *Trends in Microbiology* **24**, 440–449 (2016).

174. Roossinck, M. J. The good viruses: viral mutualistic symbioses. *Nature Reviews Microbiology* **9**, 99–108 (2011).

175. Roossinck, M. J. Move Over, Bacteria! Viruses Make Their Mark as Mutualistic Microbial Symbionts. *Journal of Virology* **89**, 6532–6535 (2015).

176. Haig, D. Retroviruses and the Placenta. *Current Biology* **22**, R609–R613 (2012).

177. Lewis, R. M., Cleal, J. K. & Hanson, M. A. Review: Placenta, evolution and lifelong health. *Placenta* **33**, S28–S32 (2012).

178. Bao, X. & Roossinck, M. J. A life history view of mutualistic viral symbioses: quantity or quality for cooperation? *Current Opinion in Microbiology* **16**, 514–518 (2013).

179. Iskra-Caruana, M. L., Baurens, F. C., Gayral, P. & Chabannes, M. A Four-Partner Plant–Virus Interaction: Enemies Can Also Come from Within. *Mol Plant Microbe Interact* **23**, 1394–1402 (2010).

180. Weise, L. I., Heymann, M., Mayr, V. & Mutschler, H. Cell-free expression of RNA encoded genes using MS2 replicase. *Nucleic Acids Res* **47**, 10956–10967 (2019).

181. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19**, 751–755 (2001).

182. Filonov, G. S., Kam, C. W., Song, W. & Jaffrey, S. R. In-Gel Imaging of RNA Processing Using Broccoli Reveals Optimal Aptamer Expression Strategies. *Chemistry & Biology* **22**, 649–660 (2015).

183. Filonov, G. S., Moon, J. D., Svensen, N. & Jaffrey, S. R. Broccoli: Rapid Selection of an RNA Mimic of Green Fluorescent Protein by Fluorescence-Based Selection and Directed Evolution. *J Am Chem Soc* **136**, 16299 (2014).

184. Katanaev, V. L., Spirin, A. S., Reuss, M. & Siemann, M. Formation of bacteriophage MS2 infectious units in a cell-free translation system. *FEBS Letters* **397**, 143–148 (1996).

185. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).

186. Tuerk, C. & Gold, L. Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase. *Science (1979)* **249**, 505–510 (1990).

187. Klug, S. J. & Famulok, M. All you wanted to know about SELEX. *Molecular Biology Reports* **20**, 97–107 (1994).

188. Shepherd, T. R. *et al.* De novodesign and synthesis of a 30-cistron translation-factor module. *Nucleic Acids Research* **45**, 10895–10905 (2017).

189. Tu, Q. *et al.* Room temperature electrocompetent bacterial cells improve DNA transformation and recombineering efficiency. *Scientific Reports* **6**, 1–8 (2016).

190. Kibbe, W. A. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* **35**, W43–W46 (2007).

191. Gasteiger, E. *et al. Protein Identification and Analysis Tools on the ExPASy Server*. *The Proteomics Protocols Handbook* vol. 1 (Humana Press, 2005).

192. Pezo, V. & Wain-Hobson, S. Hypermutagenic in vitro transcription employing biased NTP pools and manganese cations. *Gene* **186**, 67–72 (1997).

193. Garamella, J., Marshall, R., Rustad, M. & Noireaux, V. The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology. *ACS Synthetic Biology* **5**, 344–355 (2016).

194. Alam, K. K., Tawiah, K. D., Lichte, M. F., Porciani, D. & Burke, D. H. A Fluorescent Split Aptamer for Visualizing RNA-RNA Assembly in Vivo. *ACS Synthetic Biology* **6**, 1710–1721 (2017).

195. Zhuang, F., Karberg, M., Perutka, J. & Lambowitz, A. M. EcI5, a group IIB intron with high retrohoming frequency: DNA target site recognition and use in gene targeting. *RNA* **15**, 432–449 (2009).

196. Makino, K. *et al.* Complete Nucleotide Sequences of 93-kb and 3.3-kb Plasmids of an Enterohemorrhagic Escherichia coli O157:H7 Derived from Sakai Outbreak. *DNA Research* **5**, 1–9 (1998).

197. Perutka, J., Wang, W., Goerlitz, D. & Lambowitz, A. M. Use of Computer-designed Group II Introns to Disrupt Escherichia coli DExH/D-box Protein and DNA Helicase Genes. *Journal of Molecular Biology* **336**, 421–439 (2004).

198. Codon usage table. https://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=83333&aa=1&style=N.

199. Domingo-Calap, P., Cuevas, J. M. & Sanjuán, R. The Fitness Effects of Random Mutations in Single-Stranded DNA and RNA Bacteriophages. *PLOS Genetics* **5**, e1000742 (2009).

200. Friedman, S. D., Genthner, F. J., Gentry, J., Sobsey, M. D. & Vinjé, J. Gene Mapping and Phylogenetic Analysis of the Complete Genome from 30 Single-Stranded RNA Male-Specific Coliphages (Family Leviviridae) . *Journal of Virology* **83**, 11233–11243 (2009).

201. Schmidt, T. G. M. *et al.* Development of the Twin-Strep-tag® and its application for purification of recombinant proteins from cell culture supernatants. *Protein Expression and Purification* **92**, 54–61 (2013).

202. Chong, S. *et al.* Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Research* **26**, 5109–5115 (1998).

203. LeCuyer, K. A., Behlen, L. S. & Uhlenbeck, O. C. Mutants of the Bacteriophage MS2 Coat Protein That Alter Its Cooperative Binding to RNA. *Biochemistry* **34**, 10600–10606 (1995).

204. Inokuchi, Y., Kajitani, M. & Hirashima, A. A Study on the Function of the Glycine Residue in the YGDD Motif of the RNA-Dependent RNA Polymerase β-Subunit from RNA Coliphage Qβ. *The Journal of Biochemistry* **116**, 1275–1280 (1994).

205. Haapa, S., Taira, S., Heikkinen, E. & Savilahti, H. An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. *Nucleic Acids Research* **27**, 2777 (1999).

206. Craigie, R. & Mizuuchi, K. Transposition of Mu DNA: Joining of Mu to target DNA can be uncoupled from cleavage at the ends of Mu. *Cell* **51**, 493–501 (1987).

207. Rasila, T. S. *et al.* Mu transpososome activity-profiling yields hyperactive MuA variants for highly efficient genetic and genome engineering. *Nucleic Acids Research* **46**, 4649–4661 (2018).

208. Schrödinger, L., & DeLano, W. (2020). PyMOL 2.6.0a0. Retrieved from http://www.pymol.org/pymol.

209. Wilson, D. S. & Keefe, A. D. Random Mutagenesis by PCR. *Current Protocols in Molecular Biology* **51**, 8.3.1-8.3.9 (2000).

210. Cadwell, R. C. & Joyce, G. F. Randomization of Genes by PCR Mutagenesis. *PCR Methods and Applications* **2**, 28–33 (1992).

211. Smith, D., Zhong, J., Matsuura, M., Lambowitz, A. M. & Belfort, M. Recruitment of host functions suggests a repair pathway for late steps in group II intron retrohoming. *Genes & Development* **19**, 2477 (2005).

212. Cousineau, B. *et al.* Retrohoming of a Bacterial Group II Intron: Mobility via Complete Reverse Splicing, Independent of Homologous DNA Recombination. *Cell* **94**, 451–462 (1998).

213. Eskes, R., Yang, J., Lambowitz, A. M. & Perlman, P. S. Mobility of Yeast Mitochondrial Group II Introns: Engineering a New Site Specificity and Retrohoming via Full Reverse Splicing. *Cell* **88**, 865–874 (1997).

214. Qu, G., Piazza, C. L., Smith, D. & Belfort, M. Group II intron inhibits conjugative relaxase expression in bacteria by mRNA targeting. *Elife* **7**, (2018).

215. Monachello, D., Lauraine, M., Gillot, S., Michel, F. & Costa, M. A new RNA–DNA interaction required for integration of group II intron retrotransposons into DNA targets. *Nucleic Acids Research* **49**, 12394–12410 (2021).

216. Dvorak, P. *et al.* Exacerbation of substrate toxicity by IPTG in Escherichia coli BL21(DE3) carrying a synthetic metabolic pathway. *Microbial Cell Factories* **14**, 201 (2015).

217. Recorbet, G. *et al.* Conditional suicide system of Escherichia coli released into soil that uses the Bacillus subtilis sacB gene. *Appl Environ Microbiol* **59**, 1361–6 (1993).

218. Pelicic, V., Reyrat, J.-M. & Gicquel, B. Expression of the Bacillus subtilis sacB Gene Confers Sucrose Sensitivity on Mycobacteria. *Journal of Bacteriology* **178**, 1197–1199 (1996).

219. Gay, P., le Cqq, D., Steinmetz, M., Berkelman, T. & Kadol, C. I. Positive Selection Procedure for Entrapment of Insertion Sequence Elements in Gram-Negative Bacteria. *Journal of Bacteriology* **164**, 918–921 (1985).

220. Murphy, K. C. & Campellone, K. G. Lambda Red-mediated recombinogenic engineering of enterohemorrhagic and enteropathogenic E. coli. *BMC Molecular Biology* **4**, (2003).

221. Muyrers, J. P. P., Zhang, Y., Testa, G. & Stewart, A. F. Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Research* **27**, (1999).

222. Yu, D. *et al.* An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97**, 5978–5983 (2000).

223. Muyrers, J. P. P., Zhang, Y., Buchholz, F. & Stewart, A. F. RecE/RecT and Red/Red initiate double-stranded break repair by specifically interacting with their respective partners. *Genes & Development* **14**, 1971–1982 (2000).

224. Court, D. L. *et al.* RNase III: Genetics and Function; Structure and Mechanism. *Annu Rev Genet* **47**, 405 (2013).

225. Augustin, S., Müller, M. W. & Schweyen, R. J. Reverse self-splicing of group II intron RNAs in vitro. *Nature* **343**, 383–386 (1990).

226. Steitz, J. A. Isolation of the A protein from bacteriophage R17. *Journal of Molecular Biology* **33**, 937–945 (1968).

227. Rumnieks, J. & Tars, K. Crystal Structure of the Maturation Protein from Bacteriophage Qβ. *Journal of Molecular Biology* **429**, 688–696 (2017).

228. Fox, J. D. & Waugh, D. S. Maltose-Binding Protein as a Solubility Enhancer. *Methods in molecular biology* **205**, 99–117 (2003).

229. Sun, P., Tropea, J. E. & Waugh, D. S. Enhancing the Solubility of Recombinant Proteins in Escherichia coli by Using Hexahistidine-Tagged Maltose-Binding Protein as a Fusion Partner. *Methods in Molecular Biology* **705**, 259–274 (2011).

230. Nallamsetty, S. & Waugh, D. S. Solubility-enhancing proteins MBP and NusA play a passive role in the folding of their fusion partners. *Protein Expression and Purification* **45**, 175–182 (2006).

231. Hlodan, R., Craig, S. & Pain, R. H. Protein Folding and its Implications for the Production of Recombinant Proteins. *Biotechnology and Genetic Engineering Reviews* **9**, 47–88 (1991).

232. Lowry, O. *et al.* A Kinetic Study of the Competition between Renaturation and Aggregation during the Refolding of Denatured-Reduced Egg White Lysozyme. *Arch. Biochem. Biophys* **30**, 2790–2797 (1991).

233. Ventura, S. & Villaverde, A. Protein quality in bacterial inclusion bodies. *Trends in Biotechnology* **24**, 179–185 (2006).

234. Clark, E. D. B. Protein refolding for industrial processes. *Current Opinion in Biotechnology* **12**, 202–207 (2001).

235. Middelberg, A. P. J. Preparative protein refolding. *Trends in Biotechnology* **20**, 437–443 (2002).

236. Vallejo, L. F. & Rinas, U. Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microbial Cell Factories* **3**, 1–12 (2004).

237. Varshavsky, A. The N-end rule pathway and regulation by proteolysis. *Protein Science : A Publication of the Protein Society* **20**, 1298 (2011).

238. Bartel, B., Wünning, I. & Varshavsky, A. The recognition component of the N-end rule pathway. *The EMBO Journal* **9**, 3179–3189 (1990).

239. Mogk, A., Schmidt, R. & Bukau, B. The N-end rule pathway for regulated proteolysis: prokaryotic and eukaryotic strategies. *Trends in Cell Biology* **17**, 165–172 (2007).

240. Schuenemann, V. J. *et al.* Structural basis of N-end rule substrate recognition in Escherichia coli by the ClpAP adaptor protein ClpS. *EMBO Reports* **10**, 508–514 (2009).

241. Stevens, R., Stevens, L. & Price, N. The stabilities of various thiol compounds used in protein purifications. *Biochemical Education* **11**, 70 (1983).

242. Fenton, W. A., Kashi, Y., Furtak, K. & Norwich, A. L. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* **371**, 614–619 (1994).

243. Thain, A., Gaston, K., Jenkins, O. & Clarke, A. R. A method for the separation of GST fusion proteins from co-purifying GroEL. *Trends in Genetics* **12**, 209–210 (1996).

244. Rohman, M. & Harrison-Lavoie, K. J. Separation of Copurifying GroEL from Glutathione-S-Transferase Fusion Proteins. *Protein Expression and Purification* **20**, 45–47 (2000).

245. Fiedler, S. & Wirth, R. Transformation of bacteria with plasmid DNA by electroporation. *Analytical Biochemistry* **170**, 38–44 (1988).

246. Chassy, B. M., Mercenier, A. & Flickinger, J. Transformation of bacteria by electroporation. *Trends in Biotechnology* **6**, 303–309 (1988).

247. Liu, H. *et al.* Fast and efficient genetic transformation of oleaginous yeast Rhodosporidium toruloides by using electroporation. *FEMS Yeast Research* **17**, 17 (2017).

248. Suga, M. & Hatakeyma, T. High-efficiency electroporation by freezing intact yeast cells with addition of calcium. *Current Genetics* **43**, 206–211 (2003).

249. Chu, G., Hayakawa, H. & Berg, P. Electroporation for the efficient transfection of mammalian cells with DNA. *Nucleic Acids Research* **15**, 1311–1326 (1987).

250. Gehl, J. Electroporation: theory and methods, perspectives for drug delivery, gene therapy and research. *Acta Physiologica Scandinavica* **177**, 437–447 (2003).

251. Everett, J. G. & Gallie, D. R. RNA Delivery in Saccharomyces cerevisiae Using Electroporation. *YEAST* **8**, 1007–1008 (1992).

252. Gallie, D. R. Electroporation of RNA into Saccharomyces cerevisiae. *Methods in molecular biology* **47**, 81–91 (1995).

253. Kishida, T. *et al.* Sequence-specific gene silencing in murine muscle induced by electroporation-mediated transfer of short interfering RNA. *The Journal of Gene Medicine* **6**, 105–110 (2004).

254. Taketo, A. RNA transfection of Escherichia coli by electroporation. *BBA - Gene Structure and Expression* **1007**, 127–129 (1989).

255. Fouts, D. E., True, H. L. & Celander, D. W. Functional recognition of fragmented operator sites by R17/MS2 coat protein, a translational repressor. *Nucleic Acids Research* **25**, 4464–4473 (1997).

256. Batey, R. T. Structures of regulatory elements in mRNAs. *Current Opinion in Structural Biology* **16**, 299–306 (2006).

257. Gilbert, S. D., Rambo, R. P., van Tyne, D. & Batey, R. T. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nature Structural & Molecular Biology* **15**, 177–182 (2008).

258. Dethoff, E. A. *et al.* Pervasive tertiary structure in the dengue virus RNA genome. *Proc Natl Acad Sci U S A* **115**, 11513–11518 (2018).

259. McManus, C. J. & Graveley, B. R. RNA structure and the mechanisms of alternative splicing. *Current Opinion in Genetics & Development* **21**, 373–379 (2011).

260. de Borba, L. *et al.* Overlapping Local and Long-Range RNA-RNA Interactions Modulate Dengue Virus Genome Cyclization and Replication. *Journal of Virology* **89**, 3430–3437 (2015).

261. Nicholson, B. L. & White, K. A. Exploring the architecture of viral RNA genomes. *Current Opinion in Virology* **12**, 66–74 (2015).

262. Licis, N., van Duin, J., Balklava, Z. & Berzins, V. Long-range translational coupling in single-stranded RNA bacteriophages: An evolutionary analysis. *Nucleic Acids Research* **26**, 3242–3246 (1998).

263. Goldhaber-Gordon, I., Williams, T. L. & Baker, T. A. DNA Recognition Sites Activate MuA Transposase to Perform Transposition of Non-Mu DNA. *Journal of Biological Chemistry* **277**, 7694–7702 (2002).

264. Goldhaber-Gordon, I., Early, M. H. & Baker, T. A. MuA Transposase Separates DNA Sequence Recognition from Catalysis. *Biochemistry* **42**, 14633–14642 (2003).

265. Fuller, J. R. & Rice, P. A. Target DNA bending by the Mu transpososome promotes careful transposition and prevents its reversal. *Elife* (2017) doi:10.7554/eLife.21777.001.

266. Tomita, K., Ichihashi, N. & Yomo, T. Replication of partial double-stranded RNAs by Qβ replicase. *Biochemical and Biophysical Research Communications* **467**, 293–296 (2015).

267. Chetverin, A. B. Thirty Years of Studies of Qβ Replicase: What Have We Learned and What Is Yet to Be Learned? *Biochemistry* **83**, S19–S32 (2018).

268. Blumenthal, T. & Carmichael, G. G. RNA Replication: Function and Structure of Qbeta-Replicase. *Annual Review of Biochemistry* **48**, 525–548 (1979).

269. Munishkin, A. v. *et al.* Efficient templates for Qβ replicase are formed by recombination from heterologous sequences. *Journal of Molecular Biology* **221**, 463–472 (1991).

270. Roy, H. & Ibba, M. Phenylalanyl-tRNA synthetase contains a dispensable RNA-binding domain that contributes to the editing of noncognate aminoacyl-tRNA. *Biochemistry* **45**, 9156–9162 (2006).

271. Shepard, A., Shiba, K. & Schimmel, P. RNA binding determinant in some class I tRNA synthetases identified by alignment-guided mutagenesis. *Proc Natl Acad Sci U S A* **89**, 9964–9968 (1992).

272. Moras, D. & Poterszman, A. RNA-Protein Interactions: Diverse modes of recognition. *Current Biology* **5**, 249–251 (1995).

273. Beuning, P. J. *et al.* Efficient aminoacylation of the tRNA-Ala acceptor stem: Dependence on the 2:71 base pair. *RNA* **8**, 659–670 (2002).

274. Mayer, C. & RajBhandary, U. L. Conformational change of Escherichia coli initiator methionyl-tRNA fMet upon binding to methionyl-tRNA formyl transferase. *Nucleic Acids Research* **30**, 2844–2850 (2002).

275. Dahlquist, K. D. & Puglisi, J. D. Interaction of translation initiation factor IF1 with the E. coli ribosomal A site. *Journal of Molecular Biology* **299**, 1–15 (2000).

276. Petrelli, D. *et al.* Translation initiation factor IF3: two domains, five functions, one mechanism? *The EMBO Journal* **20**, 4560–4569 (2001).

277. Hall, T. C. Transfer RNA-like Structures in Viral Genomes. *International Review of Cytology* **60**, 1–26 (1979).

278. de Wachter, R., Merregaert, J., Vandenberghe, A., Contreras, R. & Fiers, W. Studies on the Bacteriophage MS2. *European Journal of Biochemistry* **22**, 400–414 (1971).

279. Croitoru, V. *et al.* RNA chaperone activity of translation initiation factor IF1. *Biochimie* **88**, 1875–1882 (2006).

280. Phadtare, S. *et al.* Transcription Antitermination by Translation Initiation Factor IF1. *Journal of Bacteriology* **189**, 4087 (2007).

281. Fowler, M. J. F. & Szekely, M. The specific interaction of Eschericia coli initiation factor IF3 with coliphage-MS2 ribonucleic acid. *Biochem Soc Trans* **7**, 979–980 (1979).

282. Johnson, B. & Szekely, M. Specific binding site of E. coli initiation factor 3 (IF3) at a 3'-terminal region of MS2 RNA. *Nature* **267**, 550–552 (1977).

283. Johnson, B. & Szekely, M. The binding site of IF-3 in MS2 RNA. *Methods Enzymol* **60**, 343–350 (1979).

284. Mills, D. R., Peterson, R. L. & Spiegelman, S. An Extracellular Darwinian Experiment with a Self-Duplicating Nucleic Acid Molecule. *PNAS* **58**, 217–224 (1967).

285. Furubayashi, T. *et al.* Emergence and diversification of a host-parasite RNA ecosystem through Darwinian evolution. *Elife* **9**, 1–15 (2020).

286. Biebricher, C. K. & Luce, R. In vitro recombination and terminal elongation of RNA by Q beta replicase. *The EMBO Journal* **11**, 5129–5135 (1992).

287. Mizuuchi, R., Furubayashi, T. & Ichihashi, N. Evolutionary transition from a single RNA replicator to a multiple replicator network. *Nature Communications* **13**, 1–10 (2022).

288. Chetverin, A. B., Chetverina, H. v. & Munishkin, A. v. On the nature of spontaneous RNA synthesis by Qβ replicase. *Journal of Molecular Biology* **222**, 3–9 (1991).

289. Ellington, A. D., Chen, X., Robertson, M. & Syrett, A. Evolutionary origins and directed evolution of RNA. *The International Journal of Biochemistry & Cell Biology* **41**, 254–265 (2009).

290. Gruber, A. R., Lorenz, R., Bernhart, S. H., Ck, R. N. & Hofacker, I. L. The Vienna RNA Websuite. *Nucleic Acids Research* **36**, (2008).

291. Lorenz, R. *et al.* ViennaRNA Package 2.0. *Algorithms for Molecular Biology : AMB* **6**, 26 (2011).

292. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189 (2009).

293. Maguin, E., Duwat, P., Hege, T., Ehrlich, D. & Gruss, A. New thermosensitive plasmid for gram-positive bacteria. *Journal of Bacteriology* **174**, 5633 (1992).

294. Krahn, P. M., O'Callaghan, R. J. & Paranchych, W. Stages in phage R17 infection: VI. Injection of a protein and RNA into the host cell. *Virology* **47**, 628–637 (1972).

295. Kozak, M. & Nathans, D. Fate of Maturation Protein during Infection by Coliphage MS2. *Nature New Biology* **234**, 209–211 (1971).

296. Lin, T., Cavarelli, J. & Johnson, J. E. Evidence for assembly-dependent folding of protein and RNA in an icosahedral virus. *Virology* **314**, 26–33 (2003).

297. Williamson, J. R. Cooperativity in macromolecular assembly. *Nature Chemical Biology* **4**, 458–465 (2008).

298. Zhang, H. X., Zhang, Y. & Yin, H. Genome Editing with mRNA Encoding ZFN, TALEN, and Cas9. *Molecular Therapy* **27**, 735–746 (2019).

299. Liu, J. *et al.* Fast and Efficient CRISPR/Cas9 Genome Editing In Vivo Enabled by Bioreducible Lipid and Messenger RNA Nanoparticles. *Advanced Materials* **31**, 1902575 (2019).

300. Aizawa, Y., Xiang, Q., Lambowitz, A. M. & Pyle, A. M. The Pathway for DNA Recognition and RNA Integration by a Group II Intron Retrotransposon. *Molecular Cell* **11**, 795–805 (2003).

301. Andersson, M. *et al.* Genome editing in potato via CRISPR-Cas9 ribonucleoprotein delivery. *Physiologia Plantarum* **164**, 378–384 (2018).

302. Gustafsson, O. *et al.* Efficient Peptide-Mediated In Vitro Delivery of Cas9 RNP. *Pharmaceutics* **13**, 878 (2021).

303. Liang, X. *et al.* Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection. *Journal of Biotechnology* **208**, 44–53 (2015).

304. Lu, J. *et al.* Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Communication and Signaling* **19**, 1–10 (2021).

305. Opuu, V., Silvert, M. & Simonson, T. Computational design of fully overlapping coding schemes for protein pairs and triplets. *Scientific Reports* **7**, 1–10 (2017).

306. Lebre, S. & Gascuel, O. The combinatorics of overlapping genes. *Journal of Theoretical Biology* **415**, 90–101 (2017).

307. Wright, B. W., Molloy, M. P. & Jaschke, P. R. Overlapping genes in natural and engineered genomes. *Nature Reviews Genetics* **23**, 154–168 (2021).

308. Groher, F. & Suess, B. Synthetic riboswitches — A tool comes of age. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* **1839**, 964–973 (2014).

309. Wittmann, A. & Suess, B. Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS Letters* **586**, 2076–2083 (2012).

310. Wahba, A. J. *et al.* Subunit I of Qβ Replicase and 30 S Ribosomal Protein Sl of Escherichia coli: Evidence for the identity of the two proteins. *Journal of Biological Chemistry* **249**, 3314–3316 (1974).

311. Schuppli, D. *et al.* Altered 3'-terminal RNA structure in phage Qbeta adapted to host factor-less Escherichia coli. *Proc Natl Acad Sci U S A* **94**, 10239–10242 (1997).

312. Su, Q., Schuppli, D., Tsui, H. C. T., Winkler, M. E. & Weber, H. Strongly Reduced Phage Qβ Replication, but Normal Phage MS2 Replication in anEscherichia coliK12 Mutant with Inactivated Qβ Host Factor (hfq) Gene. *Virology* **227**, 211–214 (1997).

313. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389–3402 (1997).

314. Altschul, S. F. *et al.* Protein Database Searches Using Compositionally Adjusted Substitution Matrices. *FEBS J* **272**, 5101 (2005).

315. Yonesaki, T., Furuse, K., Haruna, I. & Watanabe, I. Relationships among four groups of RNA coliphages based on the template specificity of GA replicase. *Virology* **116**, 379–381 (1982).

316. Haruna, I. & Spiegelman, S. Recognition of Size and Sequence by an RNA Replicase. *Biochem. Biophys. Res. Commun* **54**, 1189–1193 (1965).

317. Voigt, C. *et al.* The OB-fold proteins of the Trypanosoma brucei editosome execute RNA-chaperone activity. *Nucleic Acids Research* **46**, 10353 (2018).

318. Hajnsdorf, E. & Boni, I. v. Multiple activities of RNA-binding proteins S1 and Hfq. *Biochimie* **94**, 1544–1553 (2012).

319. Rajkowitsch, L. *et al.* RNA chaperones, RNA annealers and RNA helicases. *RNA Biology* **4**, 118–130 (2007).

320. Kamen, R., Kondo, M., Römer, W. & Weissmann, C. Reconstitution of Qβ Replicase Lacking Subunit α with Protein-Synthesis-Interference Factor i. *European Journal of Biochemistry* **31**, 44–51 (1972).

321. Scheps, R. & Revel, M. Deficiency in Initiation Factors of Protein Synthesis in Stationary-Phase Escherichia coli. *European Journal of Biochemistry* **29**, 319–325 (1972).

322. Ricciuti, C. P. Host-Virus Interactions in Escherichia coli: Effect of Stationary Phase on Viral Release from MS2-Infected Bacteria. *Journal of Virology* 162–165 (1972).

323. Kacian, D. L., Mills, D. R., Kramer, F. R. & Spiegelman, S. A Replicating RNA Molecule Suitable for a Detailed Analysis of Extracellular Evolution and Replication. *Proc Natl Acad Sci U S A* **69**, 3038 (1972).

324. Kolakofsky, D. A short biased history of RNA viruses. *RNA* **21**, 667 (2015).

325. Engelberg, H. & Soudry, E. Ribonucleic Acid Bacteriophage Release: Requirement for Host-Controlled Protein Synthesis. *Journal of Virology* **8**, 257–264 (1971).

326. Hoffmann-Berling, H. & Mazé, R. Release of male-specific bacteriophages from surviving host bacteria. *Virology* **22**, 305–313 (1964).

327. Kita, H. *et al.* Replication of Genetic Information with Self-Encoded Replicase in Liposomes. *ChemBioChem* **9**, 2403–2410 (2008).

328. Gaut, N. J. & Adamala, K. P. Reconstituting Natural Cell Elements in Synthetic Cells. *Advanced Biology* **5**, 2000188 (2021).

329. Yao, Y. *et al.* A Direct RNA-to-RNA Replication System for Enhanced Gene Expression in Bacteria. *ACS Synthetic Biology* **8**, 1067–1078 (2019).

330. Beitz, A. M., Oakes, C. G. & Galloway, K. E. Synthetic gene circuits as tools for drug discovery. *Trends in Biotechnology* **40**, 210–225 (2022).

331. Mizuuchi, R. & Ichihashi, N. Sustainable replication and coevolution of cooperative RNAs in an artificial cell-like system. *Nature Ecology & Evolution* **2**, 1654–1660 (2018).

332. Bansho, Y., Furubayashi, T., Ichihashi, N. & Yomo, T. Host-parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. *Proc Natl Acad Sci U S A* **113**, 4045–4050 (2016).

333. Fodor, E. The RNA polymerase of influenza a virus: mechanisms of viral transcription and replication. *Acta Virologica* **57**, 113–122 (2013).

334. Pflug, A., Lukarska, M., Resa-Infante, P., Reich, S. & Cusack, S. Structural insights into RNA synthesis by the influenza virus transcription-replication machine. *Virus Research* **234**, 103–117 (2017).

335. Hao, C. *et al.* Construction of RNA nanocages by re-engineering the packaging RNA of Phi29 bacteriophage. *Nature Communications* **5**, 1–7 (2014).

336. Hendrix, R. W. Bacteriophage DNA Packaging: RNA Gears in a DNA Transport Machine. *Cell* **94**, 147–150 (1998).

337. Kim, D. *et al.* Multimeric RNAs for efficient RNA-based therapeutics and vaccines. *Journal of Controlled Release* **345**, 770–785 (2022).

338. Schumacher, M. A. Bacterial plasmid partition machinery: a minimalist approach to survival. *Curr Opin Struct Biol* **22**, 72 (2012).

339. Million-Weaver, S. & Camps, M. Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? *Plasmid* **0**, 27 (2014).

7. Appendix

7.1 Abbreviations

Table 14: Abbreviations.

a.u.	Arbitrary units
AA	Amino acid
bp	Base pair
CBD	Chitin binding domain
cDNA	Complementary DNA
CDS	Coding sequence
CP	Coat protein
C-terminus	Carboxy terminus
CV	Column volume
DNA	Deoxyribonucleic acid
dsDNA	Double stranded DNA
dsRNA	Double stranded RNA
EF	Elongation factor
EMSA	Electrophoretic monility shift assay
HGT	Horizontal gene transfer
His6	Hexahistidine-tag
IB	Inclusion body
IEP	Intron encoded protein
IF	Initiation factor
IMAC	Immobilized metal ion affinity chromatography
IVT	In vitro transcription
IVTxT	In vitro coupled translation and transcription
kb	Kilobase
LB	Lysogeny broth

LP	Lysis protein
LUCA	Last universal cellular/common ancestor
MBP	Maltose binding protein
MP	Maturation protein
mRNA	Messenger RNA
MS2rep	MS2 replicase subunit
MSRP	MS2 RNA parasite
nt	Nucleotide
N-terminus	Amino terminus
NTP	Nucleotide triphosphate
OB	Oligo binding
ORF	Open reading frame
PCR	Polymerase chain reaction
PURE	Protein synthesis using recombinant elements
RBS	Ribosome binding site
RdRP	RNA-dependent RNA-polymerase
RNA	Ribonucleic acid
RNP	Ribonucleo-protein particle
rRNA	Ribosomal RNA
S1	Ribosomal protein S1
SDS-PAGE	Polyacrylamide gel-electrophoresis with sodium dodecyl sulfate
SELEX	Systematic evolution of ligands by exponential enrichment
ssRNA	Single stranded RNA
tRNA	Transfer RNA
UTR	Untranslated region

7.2 Primer Sequences

Table 15: Primer sequences.

Fragments for genomic insertions with Red/ET system	
rnc-3‘-sacB-fw	AGGTCTGTTTCGTGTGCTGAATTGTTGACGCATTTATTTATTGGTATCGCAATTAACCCTCACTAAAGGGCG
rnc-3‘-sacB-rv	AAACTTTTTGATGTTCATGGTGAATTCCTCCTGTCTGC
rnc-5‘-sacB-fw	GAGGAATTCACCATGAACATCAAAAAGTTTGCAAAACAAGC
rnc-5‘-sacB-rv	GTTCGGACGTCCGACGATGGCAATAAATCCGCAGTAACTTTTATCGATGCTAATACGACTCACTATAGGGCTC
MuA transposon cassettes	
Tnp-fr	AACGGATTACGCCAAGCTGCAGATCTGAAGCGGCGCACGAAAAACG
EcI5 reprogramming	
EBS1+2-fw	CCAAAAGGTATGTGGTTGGTTACTCCTCTNNNNNNTAGGGGTACACGGAC
EBS1+2-rv	TTCAAGCCTGTCAGCATCTTTGGCTTGTTNNNNNNAACGACGCTTCAGC
IBS1+2-fw	NNNNNNNNNGTGCGACATGAAGTCGC
IBS1+2-rv	NNNNTTTGGCATGGGAATTCTTCTAG
IVT templates for (+)-strands	
T7plus-fw	GAAATAATACGACTCACTATAGGGTGGGACCCCTTTC
plus-rv	GGGTGGTAACTAGCCAAGC
IVT templates for (-)-strands	
T7minus-fw	CAGAAGCTAATACGACTCACTATAAGGGGTGGTAACTAGCCAATCAG
minus-rv	TGGGTGGGACCCCTTTC

Reverse Transcription	
CDSII-T_{24}VN	AAGCAGTGGTATCAACGCAGAGTTTTTTTTTTTTTTTTTTTTTTTTVN
CDSII	AAGCAGTGGTATCAACGCAGAG
TSO-CDSII	GCTAATAAGCAGTGGTATCAACGCAGAGGCAGAGTACAT(rG)(rG)(rG)
Colony PCR for RZ-X plasmids	
colPCR_1	CTTGTCTGTAAGCGGATG
colPCR_2	GATTTTTGTGATGCTCGTC

Gene specific parts are underlined.

7.3 DNA sequences

Table 16: DNA sequences.

Ampliconr	**Sequence**
MuA Transposon (empty)	GATTACGCCAAGCTGCAGATCTGAAGCGGCGCACGAAAAACGCGAA AGCGTTTCACGATAAATGCGAAAACGGATCGATCCTAGTAAGCCACG TTGTGTCTCAAAATCTCTGATGTTACATTGCACAAGATAAAAATATATC ATCATGAACAATAAAACTGTCTGCTTACATAAACAGTAATACAAGGGG TGTTATG(……)TAATCAGAATTGGTTAATTGGTTGTAACACTGGCAGA GCATTACGCTGACTTGACGGGACGGCGGCTTTGTTGAATAAATCGAA CTGGATCATCCGTTTTCGCATTTATCGTGAAACGCTTTCGCGTTTTTC GTGCGCCGCTTCAGATCTCAGCTTGGCGTAATC
EcI5(ΔORF) Intron	GGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAGAATTCCCATG CCAAANNNNNNNNNNNNGTGCGACATGAAGTCGCCTGAATAATTGT TCCAGCGGAGTTCGATTCCGTCAGGGAACCTGATGTTCCGTCATCAG TAGCCTACTGACACATGCGTCACTGGTAACGGTGGGGTGTGAAGCTG TCAGGACAATGAAACCGGATCTTCGGATCGCATGAAACCGTGAGGTT ACATGTAATCTGCCAGCATCAGGCGGAGGAGGTCTAGGCTCGGTAG CATGACTAACATATGTGAACTGCTGAAGCGTCGTTNNNNNNAACAAG CCAAAGATGCTGACAGGCTTGAACCAAAAGGTATGTGGTTGGTTACT CCTCTNNNNNNTAGGGGTACACGGACAATAAACCACCGGTGTTTTGA GCAGAGTCTAACCTACTGTTGTTATTCAGGTGGAACATGGTAAGCCC GTATCGCTGCCTCGTGAGGCAGGTTAACCGCAAGGGATACTGTTGGC GGTGCGGGTAAAAGAAGGTGGAAAAAGCGAATGCCAGTCTGTAATG GACCGGATAAGGGTTGAGCCCGGCAACATTACCCTACGCGAAAGCG GGCAGACTTCCACTGGGTCTTTCATTACGAGAGAGTTTGAGTAATCTT CCAGAAAGGAAAAGCAGATGACTGAGCAGGCTACAACCTGTAAAGGT GCGTCCTTACTTAACGGTGACTCCTGGCACAGTATCAACTGGCGTCA GTGCTATCGGGAAGTGCGGAGACTGCAAGCGCGTATCGTAAAGGCA AACGCGTCACTTGTTCATGCGAAAGGTATACATGTAGTGAAACCGGC TCATGAGAGTGGGCTTAGAAAGGCTTGAGCCGTATGCCGGGAAACTG

	GCACGTACGGTTCTTAGGGGGCGGTGATGCAGTAATGCATCACTGCT ACCCGATN
Red/ET gene cassette with *sacB* for *rnc*	AGGTCTGTTTCGTGTGCTGAATTGTTGACGCATTTATTTATTGGTATC GCAATTAACCCTCACTAAAGGGCGGCCGCGAAGTTCCTATTCTCTAG AAAGTATAGGAACTTCATTCTACCGGGTAGGGGAGGCGCTTTTCCCA AGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGGCACTTGGC GCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCACCGGT AGGCGCCAACCGGCTCCGTTCTTTGGTGGCCCCTTCGCGCCACCTT CCACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGC GTCGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAGTCTCG TGCAGATGGACAGCACCGCTGAGCAATGGAAGCGGGTAGGCCTTTG GGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGGCTCAGA GGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTC AGGGGCGGGGCGGGCGCCCGAAGGTCCTCCGGAGGCCCGGCATTC TGCACGCTTCAAAAGCGCACGTCTGCCGCGCTGTTCTCCTCTTCCTC ATCTCCGGGCCTTTCGACCTGCAGCAGCACGTGTTGACAATTAATCA TCGGCATAGTATATCGGCATAGTATAATACGACAAGGTGAGGAACTAA ACCATGGGATCGGCCATTGAACAAGATGGATTGCACGCAGGTTCTCC GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAG ACGATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGG GGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAAT GAACTGCAGGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACG GGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAA GGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTC ATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAAT GCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCAC CAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCG GTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGC GCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGCATGCCCGACGGC GAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCAT GGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGG GTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTACCCGTGATATT GCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTA CGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTC TTGACGAGTTCTTCTGATTTTATCGCAACTCTCTACTTGAGCTAACAC CGTGCGTGTTGACAATTTTACCTCTGGCGGTGATAATGGTTGCAGAC AGGAGGAATTCACCATGAACATCAAAAAGTTTGCAAAACAAGCAACAG TATTAACCTTTACTACCGCACTGCTGGCAGGAGGCGCAACTCAAGCG TTTGCGAAAGAAACGAACCAAAAGCCATATAAGGAAACATACGGCATT TCCCATATTACACGCCATGATATGCTGCAAATCCCTGAACAGCAAAAA AATGAAAAATATCAAGTTCCTGAATTCGATTCGTCCACAATTAAAAATA TCTCTTCTGCAAAAGGCCTGGACGTTTGGGACAGCTGGCCATTACAA AACGCTGACGGCACTGTCGCAAACTATCACGGCTACCACATCGTCTT TGCATTAGCCGGAGATCCTAAAAATGCGGATGACACATCGATTTACAT GTTCTATCAAAAAGTCGGCGAAACTTCTATTGACAGCTGGAAAAACGC TGGCCGCGTCTTTAAAGACAGCGACAAATTCGATGCAAATGATTCTAT CCTAAAAGACCAAACACAAGAATGGTCAGGTTCAGCCACATTTACATC TGACGGAAAAATCCGTTTATTCTACACTGATTTCTCCGGTAAACATTA CGGCAAACAAACACTGACAACTGCACAAGTTAACGTATCAGCATCAG ACAGCTCTTTGAACATCAACGGTGTAGAGGATTATAAATCAATCTTTG ACGGTGACGGAAAAACGTATCAAAATGTACAGCAGTTCATCGATGAA GGCAACTACAGCTCAGGCGACAACCATACGCTGAGAGATCCTCACTA CGTAGAAGATAAAGGCCACAAATACTTAGTATTTGAAGCAAACACTGG AACTGAAGATGGCTACCAAGGCGAAGAATCTTTATTTAACAAAGCATA

	CTATGGCAAAAGCACATCATTCTTCCGTCAAGAAAGTCAAAAACTTCT GCAAAGCGATAAAAAACGCACGGCTGAGTTAGCAAACGGCGCTCTCG GTATGATTGAGCTAAACGATGATTACACACTGAAAAAAGTGATGAAAC CGCTGATTGCATCTAACACAGTAACAGATGAAATTGAACGCGCGAAC GTCTTTAAAATGAACGGCAAATGGTACCTGTTCACTGACTCCCGCGG ATCAAAAATGACGATTGACGGCATTACGTCTAACGATATTTACATGCT TGGTTATGTTTCTAATTCTTTAACTGGCCCATACAAGCCGCTGAACAA AACTGGCCTTGTGTTAAAAATGGATCTTGATCCTAACGATGTAACCTT TACTTACTCACACTTCGCTGTACCTCAAGCGAAAGGAAACAATGTCGT GATTACAAGCTATATGACAAACAGAGGATTCTACGCAGACAAACAATC AACGTTTGCGCCAAGCTTCCTGCTGAACATCAAAGGCAAGAAAACAT CTGTTGTCAAAGACAGCATCCTTGAACAAGGACAATTAACAGTTAACA AATAATTTTGTATAGAATTTACGAACGAATTCGGTACCAATAAAAGAGC TTTATTTTCATGATCTGTGTGTTGGTTTTTGTGTGCGGCGCGGAGCCC TATAGTGAGTCGTATTA<u>GCATCGATAAAAGTTACTGCGGATTTATTGC CATCGTCGGACGTCCGAAC</u>

Target specific parts are underlined.

7.4 RNA constructs

Table 17: RNA constructs.

Combined Hairpins CH7 / CH9
AAACAUGAGGAUUACCCAUGUCGAAGACAACAAAGAAGUUCAACUCUU(<u>CA</u>)CACACA CGGUAACUAGCUGCUUGGCUAGUUACC
[F30-Bro(-)]$_{UTRs(-)}$
GGGUGGUAACUAGCCAAGCAGCUAGUUACCAAAUCGGGAGAAUCCCGGGUCCUCU CUUUAGGGGGAGGUCCCUGGGCCGAAGCCCGCCCACCUUUCGGUGGAGCCGGAC CGCUUUCGCACCCGUGCUCUUUCGAGCACACCCACCCCGUUUACGGGGGUCCCUC GGUCAGCUACCGAGGAGUUGCCAUGAAUGAUCCCGAAGGAUCAUCAGAGUAUGUG GGAGCCCACACUCUACUCGACAGAUACGAAUAUCUGGACCCGACCGUCUCCCACAU ACACAUGGCAAAAACCUCCUAGGAAUGGAAUUCCGGCUACCUACAGCGAUAGCCAU GGUAGCGUCUCGCUAAAGACAUUAAAAAUGGCAUUAGCUCGACAGGAAGUUGAGCA GGACCCCGAAAGGGGUCCCACCC
[F30-Bro(+)]$_{UTRs(+)}$
GGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUUU UAAUGUCUUUAGCGAGACGCUACCAUGGCUAUCGCUGUAGGUAGCCGGAAUUCCA UUCCUAGGAGGUUUUUGCCAUGUGUAUGUGGGAGACGGUCGGGUCCAGAUAUUC GUAUCUGUCGAGUAGAGUGUGGGCUCCCACAUACUCUGAUGAUCCUUCGGGAUCA UUCAUGGCAACUCCUCGGUAGCUGACCGAGGGACCCCCGUAAACGGGGUGGGUGU GCUCGAAAGAGCACGGGUGCGAAAGCGGUCCGGCUCCACCGAAAGGUGGGCGGG CUUCGGCCCAGGGACCUCCCCCUAAAGAGAGGACCCGGGAUUCUCCCGAUUUGGU AACUAGCUGCUUGGCUAGUUACCACCC
[F30-Bro(+)]$_{UTRs(-)}$

GGGUGGUAACUAGCCAAGCAGCUAGUUACCAAAUCGGGAGAAUCCCGGGUCCUCU
CUUUAGGGGGAGGUCCCUGGGCCGAAGCCCGCCCACCUUUCGGUGGAGCCGGAC
CGCUUUCGCACCCGUGCUCUUUCGAGCACACCCACCCCGUUUACGGGGGUCCCUC
GGUCAGCUACCGAGGAGUUGCCAUGUGUAUGUGGGAGACGGUCGGGUCCAGAUA
UUCGUAUCUGUCGAGUAGAGUGUGGGCUCCCACAUACUCUGAUGAUCCUUCGGGA
UCAUUCAUGGCAAAAACCUCCUAGGAAUGGAAUUCCGGCUACCUACAGCGAUAGCC
AUGGUAGCGUCUCGCUAAAGACAUUAAAAAUGGCAUUAGCUCGACAGGAAGUUGAG
CAGGACCCCGAAAGGGGUCCCACCC

[F30-Bro(-)]$_{MS2(+)}$

GGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUUU
UAAUGUCUUUAGCGAGACGCUACCAUGGCUAUCGCUGUAGGUAGCCGGAAUUCCA
UUCCUAGGAGGUUUGACCUGUGCGAGCUUUUAGUACCCUUGAUAGGGAGAACGAG
ACCUUCGUCCCCUCCGUUCGCGUUUACGCGGACGGUGAGACUGAAGAUAACUCAU
UCUCUUUAAAAUAUCGUUCGAACUGGACUCCCGGUCGUUUUAACUCGACUGGGGC
CAAAACGAAACAGUGGCACUACCCCUCUCCGUAUUCACGGGGGGCGUUAAGUGUC
ACAUCGAUAGAUCAAGGUGCCUACAAGCGAAGUGGGUCAUCGUGGGGUCGCCCGU
ACGAUGAGAAAGCCGGUUUCGGCUUCUCCCUCGACGCACGCUCCUGCUACAGCCU
CUUCCCUGUAAGCCAAAACUUGACUUACAUCGAAGUGCCGCAGAACGUUGCGAACC
GGGCGUCGACCGAAGUCCUGCAAAAGGUCACCCAGGGUAGUUUUAACCUUGGUGU
UGCUCUAGCAGAGGCCAGGUCGACAGCCUCACAACUCGCGACGCAAACCAUUGCG
CUCGUGAAGGCGUACACUGCCGCUCGUCGCGGUAAUUGGCGCCAGGCGCUCCGC
UACCUUGCCCUAAACGAAGAUCGAAAGUUUCGAUCAAAACACGUGGCCGGCAGGU
GGUUGGAGUUGCAGUUCGGUUGGUUACCACUAAUGAGUGAUAUCCAGGGUGCAUA
UGAGAUGCUUACGAAGGUUCACCUUCAAGAGUUUCUUCCUAUGAGAGCCGUACGU
CAGGUCGGUACUAACAUCAAGUUAGAUGGCCGUCUGUCGUAUCCAGCUGCAAACU
UCCAGACAACGUGCAACAUAUCGCGACGUAUCGUGAUAUGGUUUUACAUAAACGAU
GCACGUUUGGCAUGGUUGUCGUCUCUAGGUAUCUUGAACCCACUAGGUAUAGUGU
GGGAAAAGGUGCCUUUCUCAUUCGUUGUCGACUGGCUCCUACCUGUAGGUAACAU
GCUCGAGGGCCUUACGGCCCCCGUGGGAUGCUCCUACAUGUCAGGAACAGUUACU
GACGUAAUAACGGGUGAGUCCAUCAUAAGCGUUGACGCUCCCUACGGGUGGACUG
UGGAGAGACAGGGCACUGCUAAGGCCCAAAUCUCAGCCAUGCAUCGAGGGGUACA
AUCCGUAUGGCCAACAACUGGCGCGUACGUAAAGUCUCCUUUCUCGAUGGUCCAU
ACCUUAGAUGCGUUAGCAUUAAUCAGGCAACGGCUCUCUAGAUAGAGCCCACACUC
UACUCGACAGAUACGAAUAUCUGGACCCGACCGUCUCUAUCUAGAGGGCCCUCAAC
CGGAGUUUGAAGCAUGGCUUCUAACUUUACUCAGUUCGUUCUCGUCGACAAUGGC
GGAACUGGCGACGUGACUGUCGCCCCAAGCAACUUCGCUAACGGGGUCGCUGAAU
GGAUCAGCUCUAACUCGCGUUCACAGGCUUACAAAGUAACCUGUAGCGUUCGUCA
GAGCUCUGCGCAGAAUCGCAAAUACACCAUCAAAGUCGAGGUGCCUAAAGUGGCAA
CCCAGACUGUUGGUGGUGUAGAGCUUCCUGUAGCCGCAUGGCGUUCGUACUUAAA
UAUGGAACUAACCAUUCCAAUUUUCGCUACGAAUUCCGACUGCGAGCUUAUUGUUA
AGGCAAUGCAAGGUCUCCUAAAAGAUGGAAACCCGAUUCCCUCAGCAAUCGCAGCA
AACUCCGGCAUCUACUAAUAGACGCCGGCCAUUCAAACAUGAGGAUUACCCAUGUC
GAAGACAACAAAGAAGUUCAACUCUUUAUGUAUUGAUCUUCCUCGCGAUCUUUCUC
UCGAAAUUUACCAAUCAAUUGCUUCUGUCGCUACUGGAAGCGGUGAUCCGCACAG
UGACGACUUUACAGCAAUUGCUUACUUAAGGGACGAAUUGCUCACAAAGCAUCCGA
CCUCAGGUUCCGGUAAUGACGAGGCGACCCGUCGUACCUUAGCUAUCGCUAAGCU
ACGGGAGGCGAAUGAUCGGUGCGGUCAGAUAAAUAGAGAAGGUUUCUUACAUGAC
AAAUCCUUGUCAUGGGAUCCGGAUGUUUUACAAACCAGCAUCCGUAGCCUUAUUG
GCAACCUCCUCUCUGGCUACCGAUCGUCGUUGUUUGGGCAAUGCACGUUCUCCAA
CGGUGCCUCUAUGGGGCACAAGUUGCAGGAUGCAGCGCCUUACAAGAAGUUCGCU
GAACAAGCAACCGUUACCCCCCGCGCUCUGAGAGCGGCUCUAUUGGUCCGAGACC
AAUGUGCGCCGUGGAUCAGACACGCGGUCCGCUAUAACGAGUCAUAUGAAUUUAG

GCUCGUUGUAGGGAACGGAGUGUUUACAGUUCCGAAGAAUAAUAAAAUAGAUCGG
GCUGCCUGUAAGGAGCCUGAUAUGAAUAUGUACCUCCAGAAAGGGGUCGGUGCCU
UUAUCAGACGCCGGCUCAAAUCCGUUGGUAUAGACCUAAAUGAUCAAUCGAUCAAC
CAGCGUCUGGCUCAGCAGGGCAGCGUAGAUGGUUCGCUUGCGACGAUAGACUUAU
CGUCUGCAUCCGAUUCCAUCUCCGAUCGCCUGGUGUGGAGUUUUCUCCCACCUGA
GCUAUAUUCAUAUCUCGAUCGUAUCCGCUCACACUACGGAAUCGUAGAUGGCGAG
ACGAUACGAUGGGAACUAUUUUCCACAAUGGGAAAUGGGUUCACAUUUGAGCUAGA
GUCCAUGAUAUUCUGGGCAAUAGUCAAAGCGACCCAAAUCCAUUUUGGUAACGCCG
GAACCAUAGGCAUCUACGGGGACGAUAUUAUAUGUCCCAGUGAGAUUGCACCCCG
UGUGCUAGAGGCACUUGCCUACUACGGUUUUAAACCGAAUCUUCGUAAAACGUUC
GUGUCCGGGCUCUUUCGCGAGAGCUGCGGCGCGCACUUUUACCGUGGUGUCGAU
GUCAAACCGUUUUACAUCAAGAAACCUGUUGACAAUCUCUUCGCCCUGAUGCUGAU
AUUAAAUCGGCUACGGGGUUGGGGAGUUGUCGGAGGUAUGUCAGAUCCACGCCUC
UACAAGGUGUGGGUACGGCUCUCCUCCCAGGUGCCUUCGAUGUUCUUCGGUGGG
ACGGACCUCGCUGCCGACUACUACGUAGUCAGCCCGCCUACGGCAGUCUCGGUAU
ACACCAAGACUCCGCACGGGCGGCUGCUCGCGGAUACCCGUACCUCGGGUUUCCG
UCUUGCUCGUAUCGCUCGAGAACGCAAGUUCUUCAGCGAAAAGCACGACAGUGGU
CGCUACAUAGCGUGGUUCCAUACUGGAGGUGAAAUCACCGACAGCAUGAAGUCCG
CCGGCGUGCGCGUUAUACGCACUUCGGAGUGGCUAACGCCGGUUCCCACAUUCCC
UCAGGAGUGUGGGCCAGCGAGCUCUCCUCGGUAGCUGACCGAGGGACCCCCGUAA
ACGGGGUGGGUGUGCUCGAAAGAGCACGGGUCCGCGAAAGCGGUGGCUCCACCG
AAAGGUGGGCGGGCUUCGGCCCAGGGACCUCCCCCUAAAGAGAGGACCCGGGAUU
CUCCCGAUUUGGUAACUAGCUGCUUGGCUAGUUACCACCC

[F30-Bro(-)]$_{MS2(-)}$

GGGUGGUAACUAGCCAAGCAGCUAGUUACCAAAUCGGGAGAAUCCCGGGUCCUCU
CUUUAGGGGGAGGUCCCUGGGCCGAAGCCCGCCCACCUUUCGGUGGAGCCACCG
CUUUCGCGGACCCGUGCUCUUUCGAGCACACCCACCCCGUUUACGGGGGUCCCUC
GGUCAGCUACCGAGGAGAGCUCGCUGGCCCACACUCCUGAGGGAAUGUGGGAACC
GGCGUUAGCCACUCCGAAGUGCGUAUAACGCGCACGCCGGCGGACUUCAUGCUGU
CGGUGAUUUCACCUCCAGUAUGGAACCACGCUAUGUAGCGACCACUGUCGUGCUU
UUCGCUGAAGAACUUGCGUUCUCGAGCGAUACGAGCAAGACGGAAACCCGAGGUA
CGGGUAUCCGCGAGCAGCCGCCCGUGCGGAGUCUUGGUGUAUACCGAGACUGCC
GUAGGCGGGCUGACUACGUAGUAGUCGGCAGCGAGGUCCGUCCCACCGAAGAACA
UCGAAGGCACCUGGGAGGAGAGCCGUACCCACACCUUGUAGAGGCGUGGAUCUGA
CAUACCUCCGACAACUCCCCAACCCCGUAGCCGAUUUAAUAUCAGCAUCAGGGCGA
AGAGAUUGUCAACAGGUUUCUUGAUGUAAAACGGUUUGACAUCGACACCACGGUAA
AAGUGCGCGCCGCAGCUCUCGCGAAAGAGCCCGGACACGAACGUUUUACGAAGAU
UCGGUUUAAAACCGUAGUAGGCAAGUGCCUCUAGCACACGGGGUGCAAUCUCACU
GGGACAUAUAAUAUCGUCCCCGUAGAUGCCUAUGGUUCCGGCGUUACCAAAAUGG
AUUUGGGUCGCUUUGACUAUUGCCCAGAAUAUCAUGGACUCUAGCUCAAAUGUGA
ACCCAUUUCCCAUUGUGGAAAAUAGUUCCCAUCGUAUCGUCUCGCCAUCUACGAUU
CCGUAGUGUGAGCGGAUACGAUCGAGAUAUGAAUAUAGCUCAGGUGGGAGAAAAC
UCCACACCAGGCGAUCGGAGAUGGAAUCGGAUGCAGACGAUAAGUCUAUCGUCGC
AAGCGAACCAUCUACGCUGCCCUGCUGAGCCAGACGCUGGUUGAUCGAUUGAUCA
UUUAGGUCUAUACCAACGGAUUUGAGCCGGCGUCUGAUAAAGGCACCGACCCCUU
UCUGGAGGUACAUAUUCAUAUCAGGCUCCUUACAGGCAGCCCGAUCUAUUUUAUUA
UUCUUCGGAACUGUAAACACUCCGUUCCCUACAACGAGCCUAAAUUCAUAUGACUC
GUUAUAGCGGACCGCGUGUCUGAUCCACGGCGCACAUUGGUCUCGGACCAAUAGA
GCCGCUCUCAGAGCGCGGGGGGUAACGGUUGCUUGUUCAGCGAACUUCUUGUAA
GGCGCUGCAUCCUGCAACUUGUGCCCCAUAGAGGCACCGUUGGAGAACGUGCAUU
GCCCAAACAACGACGAUCGGUAGCCAGAGAGGAGGUUGCCAAUAAGGCUACGGAU
GCUGGUUUGUAAAACAUCCGGAUCCCAUGACAAGGAUUUGUCAUGUAAGAAACCUU
CUCUAUUUAUCUGACCGCACCGAUCAUUCGCCUCCCGUAGCUUAGCGAUAGCUAA

GGUACGACGGGUCGCCUCGUCAUUACCGGAACCUGAGGUCGGAUGCUUUGUGAG
CAAUUCGUCCCUUAAGUAAGCAAUUGCUGUAAAGUCGUCACUGUGCGGAUCACCG
CUUCCAGUAGCGACAGAAGCAAUUGAUUGGUAAAUUUCGAGAGAAAGAUCGCGAG
GAAGAUCAAUACAUAAAGAGUUGAACUUCUUUGUUGUCUUCGACAUGGGUAAUCCU
CAUGUUUGAAUGGCCGGCGUCUAUUAGUAGAUGCCGGAGUUUGCUGCGAUUGCUG
AGGGAAUCGGGUUUCCAUCUUUUAGGAGACCUUGCAUUGCCUUAACAAUAAGCUC
GCAGUCGGAAUUCGUAGCGAAAAUUGGAAUGGUUAGUUCCAUAUUUAAGUACGAAC
GCCAUGCGGCUACAGGAAGCUCUACACCACCAACAGUCUGGGUUGCCACUUUAGG
CACCUCGACUUUGAUGGUGUAUUUGCGAUUCUGCGCAGAGCUCUGACGAACGCUA
CAGGUUACUUUGUAAGCCUGUGAACGCGAGUUAGAGCUGAUCCAUUCAGCGACCC
CGUUAGCGAAGUUGCUUGGGGCGACAGUCACGUCGCCAGUUCCGCCAUUGUCGAC
GAGAACGAACUGAGUAAAGUUAGAAGCCAUGCUUCAAACUCCGGUUGAGGGCCCU
CUAGAUAGAGCCCACACUCUACUCGACAGAUACGAAUAUCUGGACCCGACCGUCUC
UAUCUAGAGAGCCGUUGCCUGAUUAAUGCUAACGCAUCUAAGGUAUGGACCAUCG
AGAAAGGAGACUUUACGUACGCGCCAGUUGUUGGCCAUACGGAUUGUACCCCUCG
AUGCAUGGCUGAGAUUUGGGCCUUAGCAGUGCCCUGUCUCUCCACAGUCCACCCG
UAGGGAGCGUCAACGCUUAUGAUGGACUCACCCGUUAUUACGUCAGUAACUGUUC
CUGACAUGUAGGAGCAUCCCACGGGGGCCGUAAGGCCCUCGAGCAUGUUACCUAC
AGGUAGGAGCCAGUCGACAACGAAUGAGAAAGGCACCUUUUCCCACACUAUACCUA
GUGGGUUCAAGAUACCUAGAGACGACAACCAUGCCAAACGUGCAUCGUUUAUGUAA
AACCAUAUCACGAUACGUCGCGAUAUGUUGCACGUUGUCUGGAAGUUUGCAGCUG
GAUACGACAGACGGCCAUCUAACUUGAUGUUAGUACCGACCUGACGUACGGCUCU
CAUAGGAAGAAACUCUUGAAGGUGAACCUUCGUAAGCAUCUCAUAUGCACCCUGGA
UAUCACUCAUUAGUGGUAACCAACCGAACUGCAACUCCAACCACCUGCCGGCCACG
UGUUUUGAUCGAAACUUUCGAUCUUCGUUUAGGGCAAGGUAGCGGAGCGCCUGGC
GCCAAUUACCGCGACGAGCGGCAGUGUACGCCUUCACGAGCGCAAUGGUUUGCGU
CGCGAGUUGUGAGGCUGUCGACCUGGCCUCUGCUAGAGCAACACCAAGGUUAAAA
CUACCCUGGGUGACCUUUUGCAGGACUUCGGUCGACGCCCGGUUCGCAACGUUCU
GCGGCACUUCGAUGUAAGUCAAGUUUUGGCUUACAGGGAAGAGGCUGUAGCAGGA
GCGUGCGUCGAGGGAGAAGCCGAAACCGGCUUUCUCAUCGUACGGGCGACCCCAC
GAUGACCCACUUCGCUUGUAGGCACCUUGAUCUAUCGAUGUGACACUUAACGCCC
CCCGUGAAUACGGAGAGGGGUAGUGCCACUGUUUCGUUUUGGCCCCAGUCGAGUU
AAAACGACCGGGAGUCCAGUUCGAACGAUAUUUUAAAGAGAAUGAGUUAUCUUCAG
UCUCACCGUCCGCGUAAACGCGAACGGAGGGGACGAAGGUCUCGUUCUCCCUAUC
AAGGGUACUAAAAGCUCGCACAGGUCAAACCUCCUAGGAAUGGAAUUCCGGCUACC
UACAGCGAUAGCCAUGGUAGCGUCUCGCUAAAGACAUUAAAAAUGGCAUUAGCUCG
ACAGGAAGUUGAGCAGGACCCCGAAAGGGGUCCCACCC

MS2(wt)

GGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUUU
UAAUGUCUUUAGCGAGACGCUACCAUGGCUAUCGCUGUAGGUAGCCGGAAUUCCA
UUCCUAGGAGGUUUGACCUGUGCGAGCUUUUAGUACCCUUGAUAGGGAGAACGAG
ACCUUCGUCCCCUCCGUUCGCGUUUACGCGGACGGUGAGACUGAAGAUAACUCAU
UCUCUUUAAAAUAUCGUUCGAACUGGACUCCCGGUCGUUUUAACUCGACUGGGGC
CAAAACGAAACAGUGGCACUACCCCUCUCCGUAUUCACGGGGGGCGUUAAGUGUC
ACAUCGAUAGAUCAAGGUGCCUACAAGCGAAGUGGGUCAUCGUGGGGUCGCCCGU
ACGAUGAGAAAGCCGGUUUCGGCUUCUCCCUCGACGCACGCUCCUGCUACAGCCU
CUUCCCUGUAAGCCAAAACUUGACUUACAUCGAAGUGCCGCAGAACGUUGCGAACC
GGGCGUCGACCGAAGUCCUGCAAAAGGUCACCCAGGGUAGUUUUAACCUUGGUGU
UGCUCUAGCAGAGGCCAGGUCGACAGCCUCACAACUCGCGACGCAAACCAUUGCG
CUCGUGAAGGCGUACACUGCCGCUCGUCGCGGUAAUUGGCGCCAGGCGCUCCGC
UACCUUGCCCUAAACGAAGAUCGAAAGUUUCGAUCAAAACACGUGGCCGGCAGGU
GGUUGGAGUUGCAGUUCGGUUGGUUACCACUAAUGAGUGAUAUCCAGGGUGCAUA
UGAGAUGCUUACGAAGGUUCACCUUCAAGAGUUUCUUCCUAUGAGAGCCGUACGU

CAGGUCGGUACUAACAUCAAGUUAGAUGGCCGUCUGUCGUAUCCAGCUGCAAACU
UCCAGACAACGUGCAACAUAUCGCGACGUAUCGUGAUAUGGUUUUACAUAAACGAU
GCACGUUUGGCAUGGUUGUCGUCUCUAGGUAUCUUGAACCCACUAGGUAUAGUGU
GGGAAAAGGUGCCUUUCUCAUUCGUUGUCGACUGGCUCCUACCUGUAGGUAACAU
GCUCGAGGGCCUUACGGCCCCCGUGGGAUGCUCCUACAUGUCAGGAACAGUUACU
GACGUAAUAACGGGUGAGUCCAUCAUAAGCGUUGACGCUCCCUACGGGUGGACUG
UGGAGAGACAGGGCACUGCUAAGGCCCAAAUCUCAGCCAUGCAUCGAGGGGUACA
AUCCGUAUGGCCAACAACUGGCGCGUACGUAAAGUCUCCUUUCUCGAUGGUCCAU
ACCUUAGAUGCGUUAGCAUUAAUCAGGCAACGGCUCUCUAGAUAGGGCCCUCAAC
CGGAGUUUGAAGCAUGGCUUCUAACUUUACUCAGUUCGUUCUCGUCGACAAUGGC
GGAACUGGCGACGUGACUGUCGCCCCAAGCAACUUCGCUAACGGGGUCGCUGAAU
GGAUCAGCUCUAACUCGCGUUCACAGGCUUACAAAGUAACCUGUAGCGUUCGUCA
GAGCUCUGCGCAGAAUCGCAAAUACACCAUCAAAGUCGAGGUGCCUAAAGUGGCAA
CCCAGACUGUUGGUGGUGUAGAGCUUCCUGUAGCCGCAUGGCGUUCGUACUUAAA
UAUGGAACUAACCAUUCCAAUUUUCGCUACGAAUUCCGACUGCGAGCUUAUUGUUA
AGGCAAUGCAAGGUCUCCUAAAAGAUGGAAACCCGAUUCCCUCAGCAAUCGCAGCA
AACUCCGGCAUCUACUAAUAGACGCCGGCCAUUCAAACAUGAGGAUUACCCAUGUC
GAAGACAACAAAGAAGUUCAACUCUUUAUGUAUUGAUCUUCCUCGCGAUCUUUCUC
UCGAAAUUUACCAAUCAAUUGCUUCUGUCGCUACUGGAAGCGGUGAUCCGCACAG
UGACGACUUUACAGCAAUUGCUUACUUAAGGGACGAAUUGCUCACAAAGCAUCCGA
CCUCAGGUUCCGGUAAUGACGAGGCGACCCGUCGUACCUUAGCUAUCGCUAAGCU
ACGGGAGGCGAAUGAUCGGUGCGGUCAGAUAAAUAGAGAAGGUUUCUUACAUGAC
AAAUCCUUGUCAUGGGAUCCGGAUGUUUUACAAACCAGCAUCCGUAGCCUUAUUG
GCAACCUCCUCUCUGGCUACCGAUCGUCGUUGUUUGGGCAAUGCACGUUCUCCAA
CGGUGCCUCUAUGGGGCACAAGUUGCAGGAUGCAGCGCCUUACAAGAAGUUCGCU
GAACAAGCAACCGUUACCCCCCGCGCUCUGAGAGCGGCUCUAUUGGUCCGAGACC
AAUGUGCGCCGUGGAUCAGACACGCGGUCCGCUAUAACGAGUCAUAUGAAUUUAG
GCUCGUUGUAGGGAACGGAGUGUUUACAGUUCCGAAGAAUAAUAAAAUAGAUCGG
GCUGCCUGUAAGGAGCCUGAUAUGAAUAUGUACCUCCAGAAAGGGGUCGGUGCCU
UUAUCAGACGCCGGCUCAAAUCCGUUGGUAUAGACCUAAAUGAUCAAUCGAUCAAC
CAGCGUCUGGCUCAGCAGGGCAGCGUAGAUGGUUCGCUUGCGACGAUAGACUUAU
CGUCUGCAUCCGAUUCCAUCUCCGAUCGCCUGGUGUGGAGUUUUCUCCCACCUGA
GCUAUAUUCAUAUCUCGAUCGUAUCCGCUCACACUACGGAAUCGUAGAUGGCGAG
ACGAUACGAUGGGAACUAUUUUCCACAAUGGGAAAUGGGUUCACAUUUGAGCUAGA
GUCCAUGAUAUUCUGGGCAAUAGUCAAAGCGACCCAAAUCCAUUUUGGUAACGCCG
GAACCAUAGGCAUCUACGGGGACGAUAUUAUAUGUCCCAGUGAGAUUGCACCCCG
UGUGCUAGAGGCACUUGCCUACUACGGUUUUAAACCGAAUCUUCGUAAAACGUUC
GUGUCCGGGCUCUUUCGCGAGAGCUGCGGCGCGCACUUUUACCGUGGUGUCGAU
GUCAAACCGUUUUACAUCAAGAAACCUGUUGACAAUCUCUUCGCCCUGAUGCUGAU
AUUAAAUCGGCUACGGGGUUGGGGAGUUGUCGGAGGUAUGUCAGAUCCACGCCUC
UACAAGGUGUGGGUACGGCUCUCCUCCCAGGUGCCUUCGAUGUUCUUCGGUGGG
ACGGACCUCGCUGCCGACUACUACGUAGUCAGCCCGCCUACGGCAGUCUCGGUAU
ACACCAAGACUCCGCACGGGCGGCUGCUCGCGGAUACCCGUACCUCGGGUUUCCG
UCUUGCUCGUAUCGCUCGAGAACGCAAGUUCUUCAGCGAAAAGCACGACAGUGGU
CGCUACAUAGCGUGGUUCCAUACUGGAGGUGAAAUCACCGACAGCAUGAAGUCCG
CCGGCGUGCGCGUUAUACGCACUUCGGAGUGGCUAACGCCGGUUCCCACAUUCCC
UCAGGAGUGUGGGCCAGCGAGCUCUCCUCGGUAGCUGACCGAGGGACCCCCGUAA
ACGGGGUGGGUGUGCUCGAAAGAGCACGGGUCCGCGAAAGCGGUGGCUCCACCG
AAAGGUGGGCGGGCUUCGGCCCAGGGACCUCCCCCUAAAGAGAGGACCCGGGAUU
CUCCCGAUUUGGUAACUAGCUGCUUGGCUAGUUACCACCC

MSRP-22 (+)

UGGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUU
UUAAUGUCUUUAGCGAGACGCUACCAUGGCUAUCGCUGUAGGUAGCCGGAAUUCC

AUUCCUAGGGCUCCACCGAAAGGUGGGCGGGGCUUCGGCCCAGGGACCUCCCCC UAAAGAGAGGACCCGGGAUUCUACCGGUUUGGUAACUAGCUGAUUGGCUAGUUAC CACCCCU
[F30-Bro(+)]$_{MSRP(+)-1.0}$
UGGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUU UUAAUGUCUUUAGCGAGACGCUACCAUGGCUAUCGCUGUAGGUAGCCGGAAUUCC AUUCCUAUUGCCAUGUGUAUGUGGGAGACGGUCGGGUCCAGAUAUUCGUAUCUGU CGAGUAGAGUGUGGGCUCCCACAUACUCUGAUGAUCCUUCGGGAUCAUUCAUGGC AAGGGCUCCACCGAAAGGUGGGCGGGGCUUCGGCCCAGGGACCUCCCCCUAAAGA GAGGACCCGGGAUUCUACCGGUUUGGUAACUAGCUGAUUGGCUAGUUACCACCCC U
[F30-Bro(-)]$_{MSRP(+)-2.0m}$
UGGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUU UUAAUGUCUUUAGCGAGACGCUACCAUGGCUAUCGCUGUAGGUAGCCGGAAUUCC AUUCCUAGGGCUCCACCGAAAGGUGGGCGGGGCUUCGGCCCAGGGAUUGCCAUG AAUGAUCCCGAAGGAUCAUCAGAGUAUGUGGGAGCCCACACUCUACUCGACAGAUA CGAAUAUCUGGACCCGACCGUCUCCCACAUACACAUGGCAAACCCGGGAUUCUACC GGUUUGGUAACUAGCUGAUUGGCUAGUUACCACCCCU
[F30-Bro(+)]$_{MSRP(+)-2.0p}$
UGGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUG CCAUGUGUAUGUGGGAGACGGUCGGGUCCAGAUAUUCGUAUCUGUCGAGUAGAGU GUGGGCUCCCACAUACUCUGAUGAUCCUUCGGGAUCAUUCAUGGCAAUGUCUUUA GCGAGACGCUACCAUGGCUAUCGCUGUAGGUAGCCGGAAUUCCAUUCCUAGGGCU CCACCGAAAGGUGGGCGGGGCUUCGGCCCAGGGACCUCCCCCUAAAGAGAGGACC CGGGAUUCUACCGGUUUGGUAACUAGCUGAUUGGCUAGUUACCACCCCU
[F30-Bro(-)]$_{MSRP(+)-3.0m}$
UGGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUU UUAAUGUCUUUAGCGAGACGCUACCAUGGCUAUCGCUGUAGGUAGCCGGAAUUCC AUUCCUAGGGCUUUGCCAUGAAUGAUCCCGAAGGAUCAUCAGAGUAUGUGGGAGC CCACACUCUACUCGACAGAUACGAAUAUCUGGACCCGACCGUCUCCCACAUACACA UGGCAAGCGGGGCUUCGGCCCAGGGACCUCCCCCUAAAGAGAGGACCCGGGAUUC UACCGGUUUGGUAACUAGCUGAUUGGCUAGUUACCACCCCU
[F30-Bro(+)]$_{MSRP(+)-3.0p}$
UGGGUGGGACCCCUUUCGGGGUCCUGCUCAACUUCCUGUCGAGCUAAUGCCAUUU UUAAUGUCUUUAGCGAGACGCUACCAUGCCAUGUGUAUGUGGGAGACGGUCGGGU CCAGAUAUUCGUAUCUGUCGAGUAGAGUGUGGGCUCCCACAUACUCUGAUGAUCC UUCGGGAUCAUUCAUGGCAUAGGUAGCCGGAAUUCCAUUCCUAGGGCUCCACCGA AAGGUGGGCGGGGCUUCGGCCCAGGGACCUCCCCCUAAAGAGAGGACCCGGGAU UCUACCGGUUUGGUAACUAGCUGAUUGGCUAGUUACCACCCCU

F30-Bro(+)
UUGCCAUGUGUAUGUGGGAGACGGUCGGGUCCAGAUAUUCGUAUCUGUCGAGUAG AGUGUGGGCUCCCACAUACUCUGAUGAUCCUUCGGGAUCAUUCAUGGCAA
F30-*zeoR*(+)
UUGCCAUGUGUAUGUGGGACAGUAAUACAAGGGGUGUUAUGGCCAAGUUGACCAG UGCCGUUCCGGUGCUCACCGCGCGCGACGUCGCCGGAGCGGUCGAGUUCUGGAC CGACCGGCUCGGGUUCUCCCGGGACUUCGUGGAGGACGACUUCGCCGGUGUGGU CCGGGACGACGUGACCCUGUUCAUCAGCGCGGUCCAGGACCAGGUGGUGCCGGA CAACACCCUGGCCUGGGUGUGGGUGCGCGGCCUGGACGAGCUGUACGCCGAGUG GUCGGAGGUCGUGUCCACGAACUUCCGGGACGCCUCCGGGCCGGCCAUGACCGA GAUCGGCGAGCAGCCGUGGGGGCGGGAGUUCGCCCUGCGCGACCCGGCCGGCAA CUGCGUGCACUUCGUGGCCGAGGAGCAGGACUAACCCACAUACUCUGAUGAUCCU UCGGGAUCAUUCAUGGCAA
F30-*rep*(+)
UUGCCAUGUGUAUGUGGGUUUUGGGCUAGCAGGAGGAAUUCACCAUGCAUCACCA UCACCAUCACUCGAAGACAACAAAGAAGUUCAACUCUUUAUGUAUUGAUCUUCCUC GCGAUCUUUCUCUCGAAAUUUACCAAUCAAUUGCUUCUGUCGCUACUGGAAGCGG UGAUCCGCACAGUGACGACUUUACAGCAAUUGCUUACUUAAGGGACGAAUUGCUCA CAAAGCAUCCGACCUUAGGUUCUGGUAAUGACGAGGCGACCCGUCGUACCUUAGC UAUCGCUAAGCUACGGGAGGCGAAUGAUCGGUGCGGUCAGAUAAAUAGAGAAGGU UUCUUACAUGACAAAUCCUUGUCAUGGGAUCCGGAUGUUUUACAAACCAGCAUCCG UAGCCUUAUUGGCAACCUCCUCUCUGGCUACCGAUCGUCGUUGUUUGGGCAAUGC ACGUUCUCCAACGGUGCCUCUAUGGGGCACAAGUUGCAGGAUGCAGCGCCCUACA AGAAGUUCGCUGAACAAGCAACCGUUACCCCCCGCGCUCUGAGAGCGGCUCUAUU GGUCCGAGACCAAUGUGCGCCGUGGAUCAGACACGCGGUCCGCUAUAACGAGUCA UAUGAGUUUAGGCUCGUUGUAGGGAACGGAGUGUUUACAGUUCCGAAGAAUAAUA AAAUAGAUCGGGCUGCCUGUAAGGAGCCUGAUAUGAAUAUGUACCUCCAGAAAGG GGUCGGUGCCUUUAUCAGACGCCGGCUCAAAUCCGUUGGUAUAGACCUGAAUGAU CAAUCGAUCAACCAGCGUCUGGCUCAGCAGGGCAGCGUAGAUGGUUCGCUUGCGA CGAUAGACUUAUCGUCUGCAUCCGAUUCCAUCUCCGAUCGCCUGGUGUGGAGUUU UCUCCCACCUGAGCUAUAUUCAUAUCUCGAUCGUAUCCGCUCACACUACGGAAUCG UAGAUGGCGAGACGAUACGAUGGGAACUAUUUUCCACAAUGGGAAAUGGGUUCAC AUUUGAGCUAGAGUCCAUGAUAUUCUGGGCAAUAGUCAAAGCGACCCAAAUCCAUU UUGGUAACGCCGGAACCAUAGGCAUCUACGGGGACGAUAUUAUAUGCCCCAGUGA GAUUGCACCCCGUGUGCUAGAGGCACUUGCCUACUACGGUUUUAAACCGAAUCUU CGCAAAACGUUCGUGUCCGGGCUCUUUCGCGAGAGCUGCGGCGCGCACUUUUACC GUGGUGUCGAUGUCAAACCGUUUUACAUCAAGAAACCUGUUGACAAUCUCUUCGC CCUGAUGCUGAUAUUAAAUCGGCUACGGGGUUGGGGAGUUGUCGGAGGUAUGUCA GAUCCACGCCUCUACAAGGUGUGGGUACGGCUCUCCUCCCAGGUGCCUUCGAUGU UCUUCGGUGGGACGGACCUCGCUGCCGACUACUACGUAGUCAGCCCGCCUACGGC AGUCUCGGUAUACACCAAGACUCCGUACGGGCGGCUGCUCGCGGAUACCCGUACC UCGGGUUUCCGUCUUGCUCGUAUCGCUCGAGAACGCAAGUUCUUCAGCGAAAAGC ACGACAGUGGUCGCUACAUAGCGUGGUUCCAUACUGGAGGUGAAAUCACCGACAG CAUGAAGUCCGCCGGCGUGCGCGUUAUACGCACUUCGGAGUGGCUAACGCCGGUU CCCACAUUCCCUCAGGAGUGUGGGCCAGCGAGCUCUCCUCGGUAACCCACAUACU CUGAUGAUCCUUCGGGAUCAUUCAUGGCAA

7.5 Code for EcI5 reprogramming algorithm & Codon shuffling

```
#https://www.kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=83333
codons={'A':[('GCG',0.38),('GCC',0.31),('GCA',0.21),('GCT',0.11)],
        'C':[('TGC',0.58),('TGT',0.42)],
        'D':[('GAT',0.65),('GAC',0.35)],
        'E':[('GAA',0.70),('GAG',0.30)],
        'F':[('TTT',0.57),('TTC',0.43)],
        'G':[('GGC',0.46),('GGT',0.29),('GGA',0.13),('GGG',0.12)],
        'H':[('CAT',0.55),('CAC',0.45)],
        'I':[('ATT',0.58),('ATC',0.35),('ATA',0.07)],
        'K':[('AAA',0.73),('AAG',0.27)],
        'L':[('CTG',0.46),('TTA',0.15),('CTT',0.12),('TTG',0.12),('CTC',0.10),('CTA',0.05)],
        'M':[('ATG',1.00)],
        'N':[('AAC',0.53),('AAT',0.47)],
        'P':[('CCG',0.55),('CCT',0.17),('CCA',0.14),('CCC',0.13)],
        'Q':[('CAG',0.70),('CAA',0.30)],
        'R':[('CGC',0.44),('CGT',0.36),('CGA',0.07),('CGG',0.07),('AGG',0.03),('AGA',0.02)],
        'S':[('AGC',0.33),('TCG',0.16),('TCA',0.15),('AGT',0.14),('TCC',0.11),('TCT',0.11)],
        'T':[('ACC',0.47),('ACG',0.24),('ACT',0.16),('ACA',0.13)],
        'V':[('GTG',0.40),('GTT',0.25),('GTC',0.18),('GTA',0.17)],
        'W':[('TGG',1.00)],
        'Y':[('TAT',0.53),('TAC',0.47)],
        '*':[('TAA',0.64),('TGA',0.36),('TAG',0.00)]}

complement = [('A','T'),('G','C'),('T','A'),('C','G')]
bases = [('A','A'),('G','G'),('T','T'),('C','C')]
unpair = [('A','G'),('G','A'),('T','C'),('C','T')]

#Nucleotide frequencies from F.ZHUANG et al. 2009, for -7 frequency is a combination of values of the selection from IBS1 & IBS2
#IEP binding -26 to -14 &  +2 to +10
#   equals     0 to +12 & +28 to +36
#   -8 to  -6 for EBS1/IBS1 & EBS2/IBS2 with -8 for no base pairing (EBS2) & -6 for  no/base pairing (5 or 6 bp EBS1)
#= +18 to +20
freq_tight={-26:[('A',0.760),('C',1.500),('G',0.900),('T',1.200)],
          -25:[('A',0.864),('C',1.000),('G',1.051),('T',1.000)],
          -24:[('A',1.036),('C',1.000),('G',0.902),('T',1.167)],
          -23:[('A',1.091),('C',0.692),('G',1.041),('T',0.941)],
          -22:[('A',1.474),('C',0.450),('G',1.106),('T',0.733)],
          -21:[('A',1.037),('C',1.125),('G',1.077),('T',0.684)],
          -20:[('A',1.100),('C',1.353),('G',0.760),('T',1.214)],
          -19:[('A',1.353),('C',0.500),('G',1.244),('T',0.579)],
          -18:[('A',0.250),('C',4.875),('G',1.133),('T',0.150)],
          -17:[('A',0.458),('C',4.923),('G',0.440),('T',0.214)],
          -16:[('A',0.667),('C',1.143),('G',1.023),('T',1.231)],
          -15:[('A',4.050),('C',0.050),('G',0.326),('T',0.200)],
          -14:[('A',3.846),('C',0.000),('G',0.000),('T',0.000)],
          -13:[('A',0.192),('C',1.857),('G',1.412),('T',0.808)],
          -12:[('A',0.750),('C',0.278),('G',1.262),('T',1.250)],
          -11:[('A',0.294),('C',0.158),('G',1.818),('T',0.600)],
          -10:[('A',0.067),('C',0.333),('G',1.933),('T',0.263)],
          -9:[('A',0.125),('C',0.100),('G',1.457),('T',1.611)],
          -8:[('A',0.632),('C',0.056),('G',1.854),('T',0.524)],
          -7:[('A',0.689),('C',1.396),('G',1.628),('T',1.714)],
          -6:[('A',0.900),('C',0.605),('G',3.308),('T',0.552)],
          -5:[('A',0.452),('C',1.710),('G',0.650),('T',1.105)],
          -4:[('A',0.968),('C',1.111),('G',0.947),('T',0.957)],
          -3:[('A',1.269),('C',0.625),('G',1.615),('T',1.000)],
          -2:[('A',1.941),('C',0.707),('G',0.833),('T',0.958)],
          -1:[('A',0.227),('C',1.571),('G',0.353),('T',1.308)],
          1:[('A',2.158),('C',1.682),('G',0.481),('T',0.313)],
          2:[('A',1.900),('C',0.525),('G',0.682),('T',1.368)],
          3:[('A',0.769),('C',1.465),('G',0.200),('T',0.800)],
          4:[('A',0.870),('C',1.091),('G',0.286),('T',1.400)],
          5:[('A',0.625),('C',0.150),('G',0.250),('T',2.759)],
```

```
          6:[('A',1.462),('C',0.646),('G',2.545),('T',0.759)],
          7:[('A',1.000),('C',1.000),('G',1.176),('T',0.864)],
          8:[('A',1.000),('C',1.158),('G',0.864),('T',0.800)],
          9:[('A',0.826),('C',1.103),('G',1.214),('T',0.840)],
          10:[('A',0.700),('C',0.905),('G',0.714),('T',2.067)]}

freq_loose={-26:[('A',0.760),('C',1.500),('G',0.900),('T',1.200)],
          -25:[('A',0.864),('C',1.000),('G',1.051),('T',1.000)],
          -24:[('A',1.036),('C',1.000),('G',0.902),('T',1.167)],
          -23:[('A',1.091),('C',0.692),('G',1.041),('T',0.941)],
          -22:[('A',1.474),('C',0.450),('G',1.106),('T',0.733)],
          -21:[('A',1.037),('C',1.125),('G',1.077),('T',0.684)],
          -20:[('A',1.100),('C',1.353),('G',0.760),('T',1.214)],
          -19:[('A',1.353),('C',0.500),('G',1.244),('T',0.579)],
          -18:[('A',0.250),('C',4.875),('G',1.133),('T',0.150)],
          -17:[('A',0.458),('C',4.923),('G',0.440),('T',0.214)],
          -16:[('A',0.667),('C',1.143),('G',1.023),('T',1.231)],
          -15:[('A',4.050),('C',0.050),('G',0.326),('T',0.200)],
          -14:[('A',3.846),('C',0.000),('G',0.000),('T',0.000)],
          -13:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          -12:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          -11:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          -10:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          -9:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          -8:[('A',0.632),('C',0.056),('G',1.854),('T',0.524)],
          -7:[('A',0.689),('C',1.396),('G',1.628),('T',1.714)],
          -6:[('A',0.900),('C',0.605),('G',3.308),('T',0.552)],
          -5:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          -4:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          -3:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          -2:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          -1:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          1:[('A',1.000),('C',1.000),('G',1.000),('T',1.000)],
          2:[('A',1.900),('C',0.525),('G',0.682),('T',1.368)],
          3:[('A',0.769),('C',1.465),('G',0.200),('T',0.800)],
          4:[('A',0.870),('C',1.091),('G',0.286),('T',1.400)],
          5:[('A',0.625),('C',0.150),('G',0.250),('T',2.759)],
          6:[('A',1.462),('C',0.646),('G',2.545),('T',0.759)],
          7:[('A',1.000),('C',1.000),('G',1.176),('T',0.864)],
          8:[('A',1.000),('C',1.158),('G',0.864),('T',0.800)],
          9:[('A',0.826),('C',1.103),('G',1.214),('T',0.840)],
          10:[('A',0.700),('C',0.905),('G',0.714),('T',2.067)]}

#Define input function
def inp_check (what,typ,max_w):
    while True:
        tm_ip = input('%s: '%what)
        try:
            typ_tm_ip = typ(tm_ip)
        except:
            print('Invalid input! Try again.')
            continue
        if (typ == int or typ == float and max_w >= typ_tm_ip >= 0):
                return(typ_tm_ip)
                break
        elif (typ == str and len(tm_ip) > 0):
            return(typ_tm_ip)
            break
        else:
            print('Invalid input! Try again.')
            continue

#Define incrementation function:
def increment (seq_in, length, move):
    seq = seq_in
    temp = list()
    while len(seq)>=length:
        try:
            temp.append(seq[:length])
```

```
            seq = seq[move:]
        except:
            break
    return(temp)

#Define translation function
def translate (seq_in, ref, inv):
    temp = ''
    if type(ref) == dict:
        for x in range(len(seq_in)):
            for key in ref:
                for y in range(len(ref[key])):
                    if inv == 'y':
                        if not seq_in[x] in ref[key][y]:
                            continue
                        else:
                            temp = key + temp
                    else:
                        if not seq_in[x] in ref[key][y]:
                            continue
                        else:
                            temp = temp + key
    elif type(ref) == list:
        for x in range(len(seq_in)):
            for y in range(len(ref)):
                if inv == 'y':
                    if not seq_in[x] in ref[y][0]:
                        continue
                    else:
                        temp = ref[y][1] + temp
                else:
                    if not seq_in[x] in ref[y][0]:
                        continue
                    else:
                        temp = temp + ref[y][1]
    return(temp)

#Define scoring method
def score(seq_in,ref,thresh):
    tscore = 0
    key = []
    for dkey in ref:
        key.append(dkey)
    slst = []
    for x in range(len(key)):
        for y in ref[key[x]]:
            if not seq_in[x] in y:
                continue
            else:
                slst.append(y[1])
    for val in slst:
        if val == 0:
            tscore = tscore -30
        else:
            tscore = tscore + math.log(val,2)
    if tscore>=thresh:
        temp_hit = (tscore,seq_in)
        return(temp_hit)

#Get sequence
while True:
    raw_seq = inp_check('Enter Sequence (>36 bp)', str,100000)
    try:
        all_char = ['A','C','G','T']
        sequence = ''
        for chara in raw_seq:
            if not chara.upper() in all_char:
                continue
            else:
```

```
            sequence = sequence + chara.upper()
        if len(sequence) >= 36:
            break
        else:
            conti = inp_check('No valid sequence! Continue [y/n]?', str,1000)
            if conti == 'y':
                continue
            else:
                sys.exit()
    except:
        print('--- Exit program. ---')
        sys.exit()

#Get date
datum_f = str(datetime.datetime.now())
date = datum_f[:4]+datum_f[5:7]+datum_f[8:10]

#Get name
name = inp_check('Sequence name',str,1000)

#Choose frequency-table
restrict = inp_check('''Loose or tight score calculation? [l/t]
    Loose: lower training bias & more hits
    Tight: potentially more accurate but less hits''',str,1000)
if restrict == 'l':
    freq = freq_loose
    sc = 'loose'
    print('Loose score calculation!')
else:
    freq = freq_tight
    sc = 'tight'
    print('Loose score calculation!')
    freq = freq_tight
    sc = 'tight'
    print('Tight score calculation!')

#Set minimum log-odds score
score_gw = inp_check('''Enter minimum score (0 to 16)
    Max score = 29 for tigth restrictions calculation
    Max score = 23 for loose restrictions calculation
    Higher scores are better, good scores >=8''', float,13)
print('Minimum score: ',score_gw)

#Define sense, antisense & coding sequence from input
sense = translate(sequence, bases, 'n')
anti = translate(sequence,complement, 'y')
strt_cod = sense.find('ATG')
cds_sense = sense[strt_cod:(len(sense)-len(sense[strt_cod:])%3)]
prot = translate(increment(cds_sense,3,3),codons,'n')
sn = sense
snl = []
while len(sn) > 0:
    snl.append(sn[:120])
    sn = sn[120:]

if len(cds_sense) >= 39:
    #Shuffle codons
    shuff = inp_check('''Shuffle codons? [y/n]
    Test for better target sequences with synonymous codons.
    Will require more time / produces significantly bigger output file!
    Increasing minimum score or restricting codons is recommended!
    Coding sequence will start at the very first ATG
    CDS does not stop at the first stop codon!''',str,1000)
    if not shuff == 'y':
        shuff = 'n'
else:
    shuff = 'n'
```

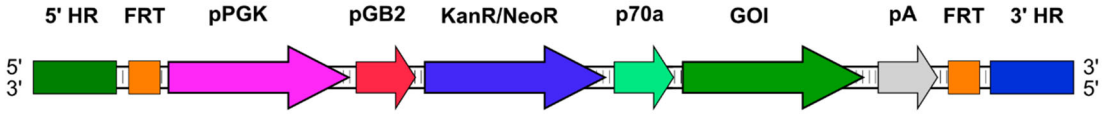
```
#Restrict codon shuffling to codons with highest fraction, reduces number of codon permutations significantly!!!
if shuff == 'y':
    codon_restric = inp_check('''Restrict to # number of codons with highest frequency?
(reduces computing requirements!) [#1-6/n]''', str, 1000)
    try:
        codon_restric = int(codon_restric)
        if 7>codon_restric>0:
            grw = codon_restric
        else:
            codon_restric = 'n'
    except:
        codon_restric = 'n'
    if codon_restric == 'n':
        cod_shu = codons
        print('Shuffling not restricted!')
    else:
        cod_res = {}
        for k in codons:
            sort = sorted(codons[k], key=lambda tup: tup[1], reverse=True)
            res = sort[:grw]
            tmp = {k : res}
            cod_res.update(tmp)
        cod_shu = cod_res
        print('Shuffling restricted.')

#Define position of insertion of Ecl5-intron
def posit(seq_in,ori,shuff):
    pos = ''
    if shuff == 'y' and ori == 's':
        prots = translate(increment(seq_in,3,3), codons, 'n')
        pos_rel = prot.find(prots)
        pos = strt_cod + pos_rel*3 + 26
    elif shuff == 'y' and ori == 'a':
        prota = translate(increment(translate(seq_in,complement,'y'),3,3), codons, 'n')
        pos_rel = prot.find(prota)
        pos = strt_cod + pos_rel*3 + 13
    elif shuff == 'n' and ori == 's':
        pos_rel = sense.find(seq_in)
        pos = pos_rel + 26
    elif shuff == 'n' and ori == 'a':
        pos_rel = anti.find(seq_in)
        pos = len(anti[pos_rel:]) - 26
    else:
        print('Error!')
    return(pos)

#Recombine codons from degenerate AA-sequence
def shuffle (seq_in):
    counter = 0
    for a in seq_in:
        title = 'tmp'+str(counter)+'.txt'
        if counter == 0:
            tmp_n = open(title,'w+')
            for AA in cod_shu:
                if not a == AA:
                    continue
                else:
                    for trip in cod_shu[AA]:
                        tmp_n.write(trip[0]+'\n')
            tmp_n.close()
        else:
            title_n = 'tmp'+str(counter-1)+'.txt'
            tmp_n = open(title_n,'r')
            tmp = open(title,'w')
            for line in tmp_n:
                nl = line.find('\n')
                line = line[:nl]
                for AA in cod_shu:
```

```
                if not a == AA:
                    continue
                else:
                    for trip in cod_shu[AA]:
                        tmp.write(line+trip[0]+'\n')
            tmp_n.close()
            tmp.close()
            os.remove(title_n)
        counter = counter + 1
    temp = open('temp.txt','w+')
    title = 'tmp'+str(counter-1)+'.txt'
    tmp = open(title,'r')
    for line in tmp:
        temp.write(line)
    tmp.close()
    temp.close()
    os.remove(title)

pr1_fw = 'gtgcgacatgaagtcgc'
pr1_rv = 'tttggcatgggaattcttctag'
pr2_fw = ('ccaaaaggtatgtggttggttactcctct','taggggtacacggac')
pr2_rv = ('ttcaagcctgtcagcatctttggcttgtt','aacgacgcttcagc')

#Define ouput format
def output(data,ori,shuffle,file):
    if shuffle == 'y':
        pos = data[0]
        dat = (data[1],data[2])
        shu = 'Yes'
    else:
        pos = posit(data[1],ori,'n')
        dat = (data[0],data[1])
        shu = 'No'
    if ori == 's':
        pos_s = pos - 25
        pos_e = pos + 10
        orient = 'sense'
    elif ori == 'a':
        pos_s = pos - 9
        pos_e = pos + 26
        orient = 'antisense'
    sq = dat[1][:26]+'|'+dat[1][26:]
    file.write('Score: %s\t Insertion Site: %s\nPosition:\t%s to %s \tinsertion after %s (sense) in %s-strand\nCodons
shuffled?\t%s\n'%(round(dat[0],2),sq,pos_s,pos_e,pos,orient,shu))
    IBS1 = dat[1][20:26]
    IBS2 = dat[1][13:19]
    IBS3 = dat[1][26]
    EBS1_s = translate(IBS1,complement,'y')
    EBS1_f = translate(IBS1[1:],complement,'y')+translate(IBS1[0],unpair,'n')
    EBS2 = translate(IBS2[-1],unpair,'n')+translate(IBS2[:-1],complement,'y')
    EBS3 = translate(IBS3,complement,'y')
    file.write('IBS1: %s \t\t\t\t\t IBS2: %s  \t\t IBS3: %s \n'%(IBS1,IBS2,IBS3))
    file.write('EBS1: %s/%s (5 bp/6 bp with IBS1)   \t EBS2: %s \t\t EBS3: %s \n'%(EBS1_f,EBS1_s,EBS2,EBS3))
    pr1f = IBS2[-2:]+IBS1+pr1_fw
    pr1r = translate(IBS2[:4],complement,'y')+pr1_rv
    pr2ff = pr2_fw[0]+EBS1_f+pr2_fw[1]
    pr2fs = pr2_fw[0]+EBS1_s+pr2_fw[1]
    pr2r = pr2_rv[0]+translate(EBS2,complement,'y')+pr2_rv[1]
    file.write('''Plasmid: EcI5_E3%s/I3%s
Primer1_fw\t %s
Primer1_rv\t %s
Primer2_fw\t %s / %s
Primer2_rv\t %s
\n'''%(EBS3,IBS3,pr1f,pr1r,pr2ff,pr2fs,pr2r))

#Change directory
dire = date+'_'+name+'_'+restrict.upper()
cwd = os.getcwd()
try:
```

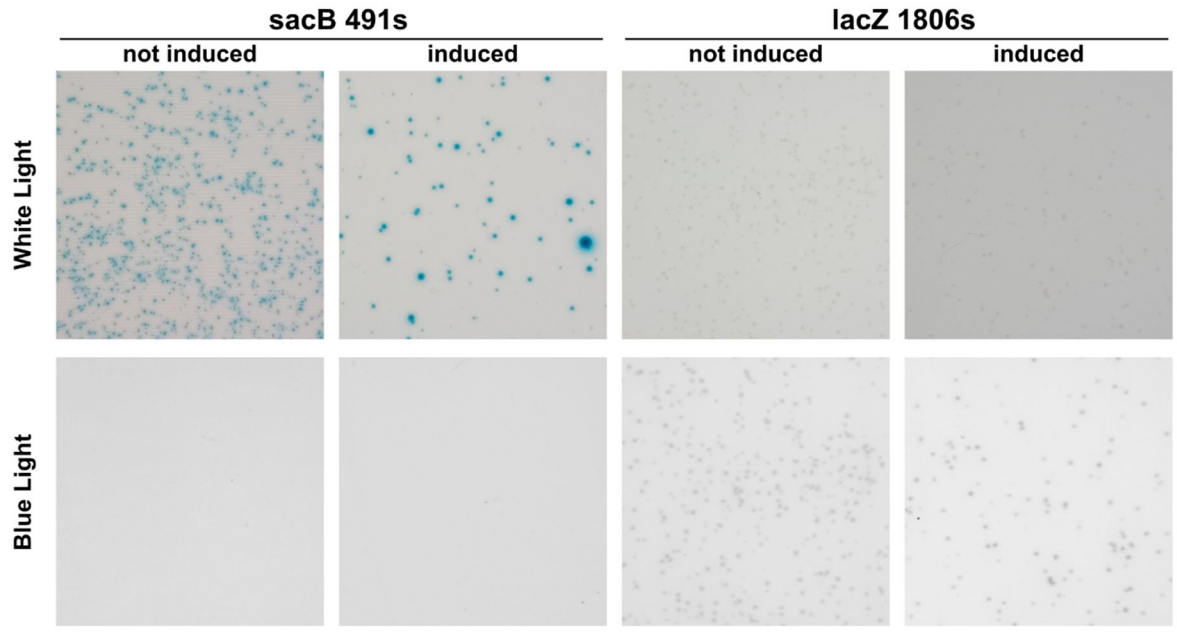
```
    path = os.path.join(cwd,dire)
    os.mkdir(path)
except:
    dire =dire+'_'+datum_f[11:13]+datum_f[14:16]+datum_f[17:19]
    path = os.path.join(cwd,dire)
    os.mkdir(path)
os.chdir(path)
print('\nOutput file created in '+os.getcwd())

#Write header for basic score calculation
score_t = str(score_gw).split('.')
score_h = score_t[0]+'-'+ score_t[1]
save = open('%s_%s_%s_%s.txt'%(date,name,score_h,sc),'w+')
save.write('''_____EcI5 Insertion site scanner_____\n
Date: %s\t\tInput: %s\t\tMinimum score: %s\t\tCalculation: %s\t\n
Input sequence:\n\n'''%(date,name,score_gw,sc))
for x in range(len(snl)):
    save.write('%i\t%s\n'%((1+x*120),snl[x]))
save.write('_____\n\n')
save.close()
#Output for sense strand:
sense_inc = increment(sense,len(freq),1)
sen_score = sorted([score(x,freq,score_gw) for x in sense_inc if score(x,freq,score_gw) != None], reverse=True)
save = open('%s_%s_%s_%s.txt'%(date,name,score_h,sc),'a+')
[output(data,'s','n',save) for data in sen_score]
save.write('\n\t----------End of sense block I----------\n\n\n')
save.close()
print('\nDone with sense block 1!')
#Output for antisense strand
anti_inc = increment(anti,len(freq),1)
ant_score = sorted([score(x,freq,score_gw) for x in anti_inc if score(x,freq,score_gw) != None], reverse=True)
save = open('%s_%s_%s_%s.txt'%(date,name,score_h,sc),'a+')
[output(data,'a','n',save) for data in ant_score]
save.write('\n\t----------End of antisense block I----------\n\n\n')
save.close()
print('\nDone with antisense block 1!')

if shuff == 'y':
    print('\nStart shuffling.')
    #Write header for shuffled score calculation
    score_t = str(score_gw).split('.')
    score_h = score_t[0]+'-'+ score_t[1]
    save = open('%s_%s_%s_%s_shuff.txt'%(date,name,score_h,sc),'w+')
    save.write('''_____EcI5 Insertion site scanner_____\n
Date: %s\t\tInput: %s\t\tMinimum score: %s\t\tCalculation: %s\n
Input sequence:\n\n'''%(date,name,score_gw,sc))
    for x in range(len(snl)):
        save.write('%i\t%s\n'%((1+x*120),snl[x]))

save.write('_____\n\n')
    save.close()
    #Shuffle & score calculation
    pept_inc = increment(prot,13,1)
    #Sense strand
    tmp_sense = open('tmp_sense.txt','w+')
    for snip in pept_inc:
        shuffle(snip)
        temp = open('temp.txt','r')
        for line in temp:
            pos_s = posit(line,'s','y')
            frag_sen = increment(line[:-1],36,1)
            for x in range(len(frag_sen)):
                frags = frag_sen[x]
                sco_sen = score(frags,freq,score_gw)
                if sco_sen != None and len(frags)==36:
```

```
                pos_s = pos_s + x
                tmp_sense.write('%i %f %s\n'%(pos_s,sco_sen[0],sco_sen[1]))
        temp.close()
        os.remove('temp.txt')
    dfs = pd.read_csv('tmp_sense.txt',' ',header=0, names=['Pos.','Score','Seq.'])
    dfsg = dfs.groupby('Pos.')
    dfsgm = dfsg.max()
    rec_s = dfsgm.to_records(index=True)
    res_s = sorted(list(rec_s), key=lambda tup: tup[1], reverse=True)
    save = open('%s_%s_%s_%s_shuff.txt'%(date,name,score_h,sc),'a+')
    [output(data,'s','y',save) for data in res_s]
    tmp_sense.close()
    save.write('\n\t------------------End of sense block II------------------\n\n\n')
    save.close()
    print('\nDone with shuffling sense block 2!')
    #Antisense strand
    tmp_anti = open('tmp_anti.txt','w+')
    for snip in pept_inc:
        shuffle(snip)
        temp = open('temp.txt','r')
        for line in temp:
            pos_a = posit(line,'s','y')
            frag_ant = increment(translate(line[:-1],complement,'y'),36,1)
            for x in range(len(frag_ant)):
                fraga = frag_ant[x]
                sco_ant = score(fraga,freq,score_gw)
                if sco_ant != None and len(fraga)==36:
                    pos_a = pos_a - x -13
                    tmp_anti.write('%i %f %s\n'%(pos_a,sco_ant[0],sco_ant[1]))
        temp.close()
        os.remove('temp.txt')
    dfa = pd.read_csv('tmp_anti.txt',' ',header=0, names=['Pos.','Score','Seq.'])
    dfag = dfa.groupby('Pos.')
    dfagm = dfag.max()
    rec_a = dfagm.to_records(index=True)
    res_a = sorted(list(rec_a), key=lambda tup: tup[1], reverse=True)
    save = open('%s_%s_%s_%s_shuff.txt'%(date,name,score_h,sc),'a+')
    [output(data,'a','y',save) for data in res_a]
    tmp_anti.close()
    save.write('\n\t------------------End of antisense block II------------------\n\n\n')
    save.close()
    print('\nDone with shuffling antisense block 2!')

print('\nDone!')
os.chdir(cwd)
sys.exit()
```

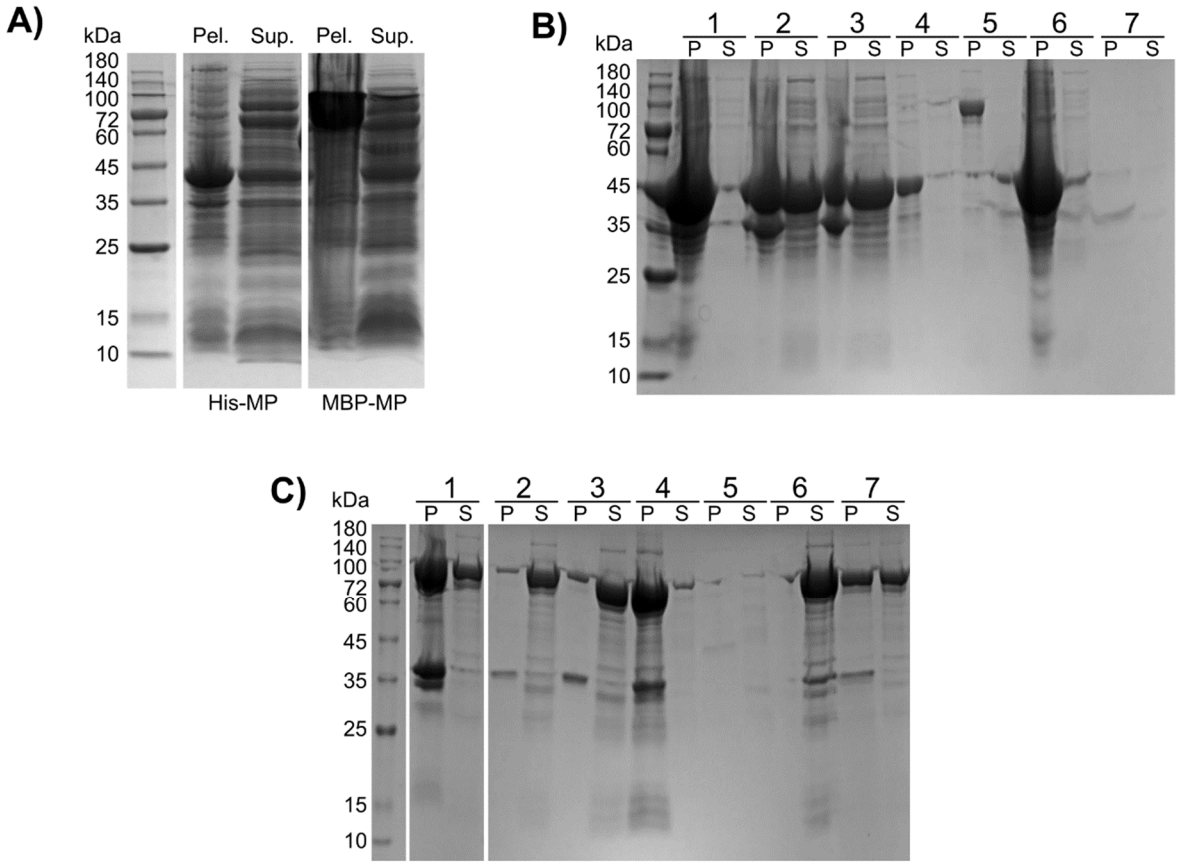
7.6 Supplementary Figures



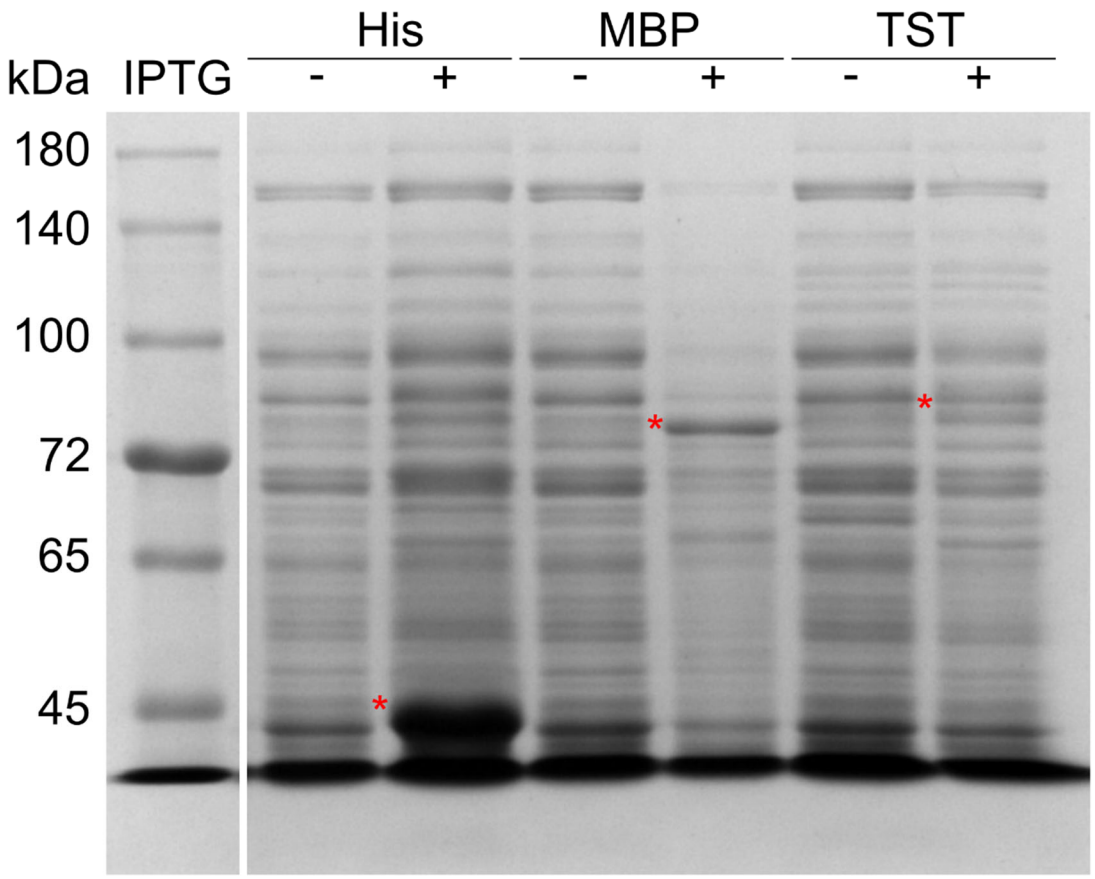
Supplementary Figure 1 Expanded gene cassette for genomic engineering: The gene cassette is flanked by the 5' and 3' homologous regions (HR), as well as recognition sites for FLP recombinase, enabling removal of the gene cassette from the target locus. Furthermore, the cassette contains both a eukaryotic (pPGK) and a prokaryotic promoter (pGB2), for the expression of the resistance against Neomycin and Kanamycin (KanR/NeoR), respectively. The selection marker is followed by the constitutive p70a promoter (p70a) for expression of a gene of interest (GOI) in prokaryotes, in this study *sacB*. Additionally, the cassette has a poly-adenylation signal (pA) for the expression in eukaryotes.



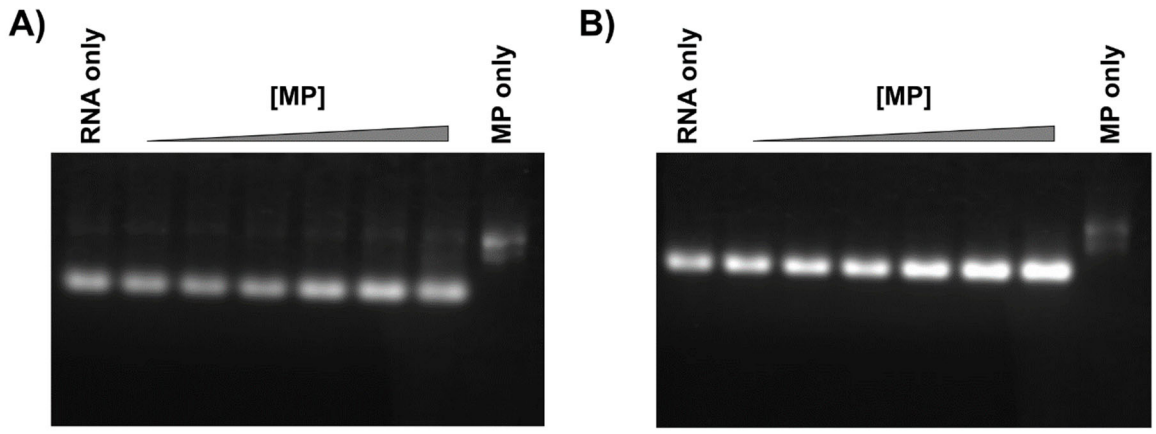
Supplementary Figure 2 Disruption of the *lacZ* gene with EcI5 intron: Images of the colonies of *E. coli* HMS-174(DE3)^{sacB} harbouring plasmids encoding for the EcI5() and IEP targeting either *sacB* or *lacZ* from at positions 491s and 1806s, respectively, subjected to the blue/white screening as shown in Figure 9. Images were taken as photographies with white light illumination, where blue colonies are well but white colonies only barely visible, or in a blue light scanner (Microtek Bio-1000F), where the visibility was inverted.



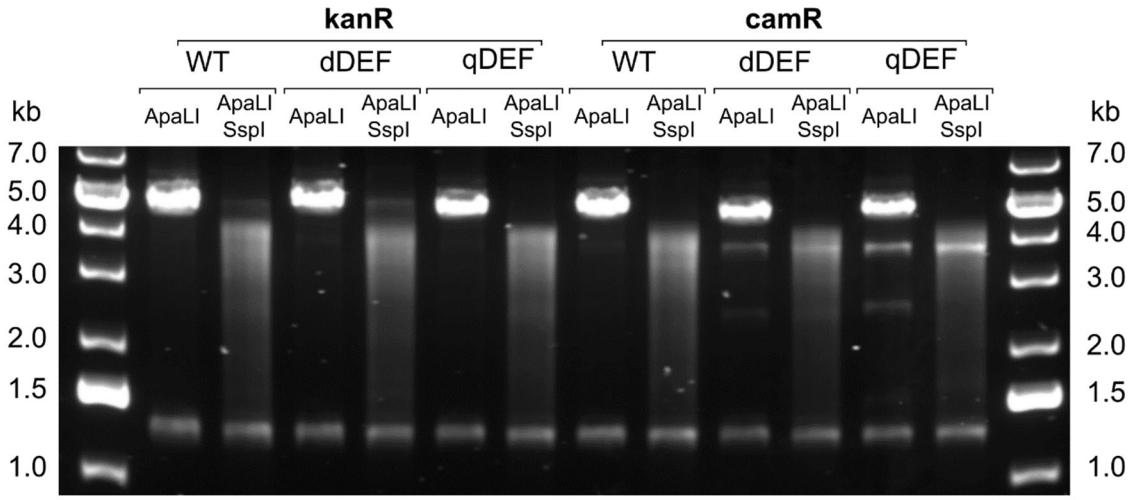
Supplementary Figure 3 Solubility of recombinantly expressed MP: A) Exemplary SDS-Page for the overexpression of N-terminally tagged MP variants His-MP (46 kDa) and MBP-MP (87 kDa). The gel shows the distribution of overexpressed MP-protein between the supernatant (Sup.) and the pelleted cell debris (Pel.) after cell lysis. **B)** and **C)** SDS-Page of samples taken from the evaluation of denaturing conditions for His-MP (**B)**) and MBP-MP (**C)**), respectively. The gels show the distribution of MP-protein between the soluble fraction (S) and the pellet fraction (P) after denaturation and centrifugation to remove aggregated protein. Numbers 1 – 7 refer to the tested conditions listed in Table 11**.**



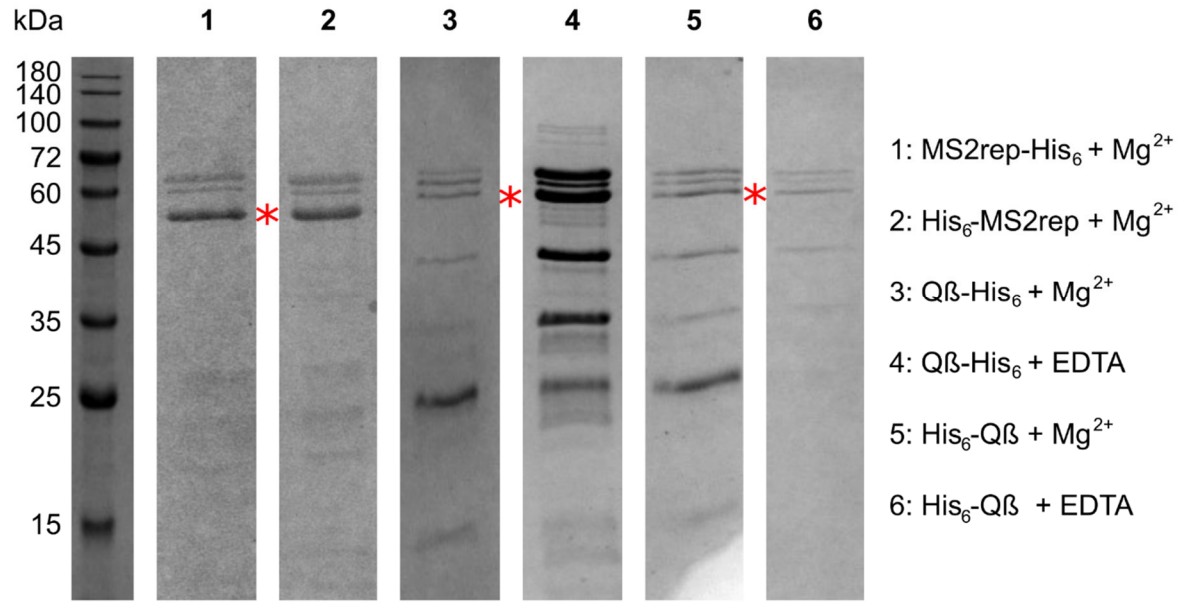
Supplementary Figure 4 Induction dependant overexpression of N-terminal mutants of MP variants: SDS-Page of samples taken from cells harbouring the expression plasmids for MP(mut)-His (His), MP(mut)-MBP (MBP) and MP(mut)-TST (TST), with or without addition of IPTG for the induction of overexpression. Red asterisks indicate the expected position of bands corresponding to the respective MP(mut) variant.



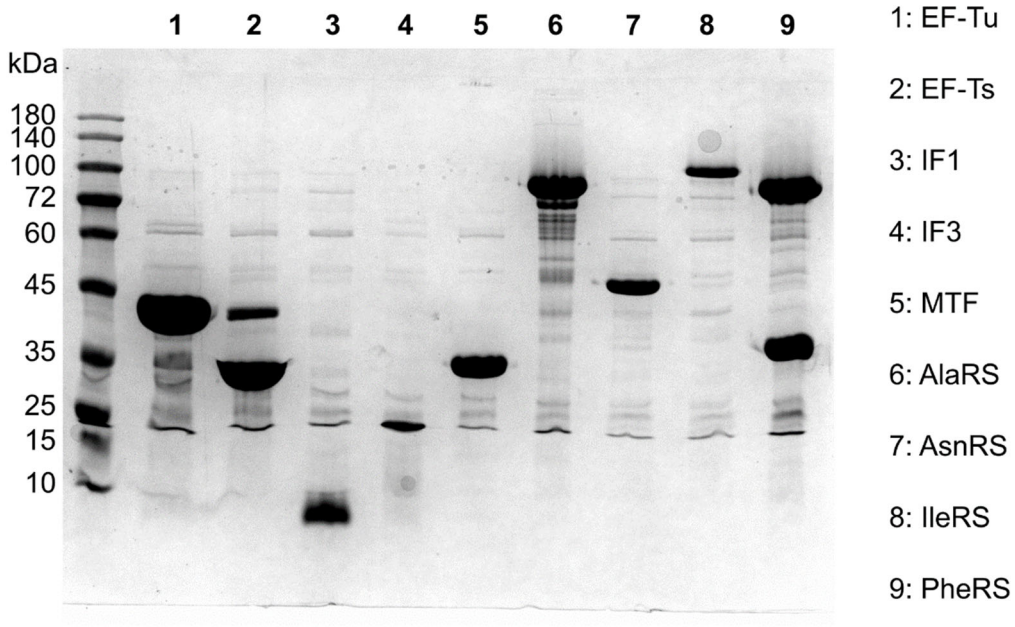
Supplementary Figure 5 EMSA of synthetic MS2-derived RNA replisomes: A) and **B)** EMSA of [F30-Bro(+)]$_{MSRP(+)-1.0}$ (**A)**) or respectively [ZeoR(+)]$_{MSRP(+)-2.0p}$ (**B)**) with varying concentrations of isolated MP-MBP (approximately 0.8 nM to 0.2 µM). Samples were prepared identically to samples for Figure 15D, but with 100 nM RNA instead of 25 nM, and run on a 1.5 % TA agarose gel.



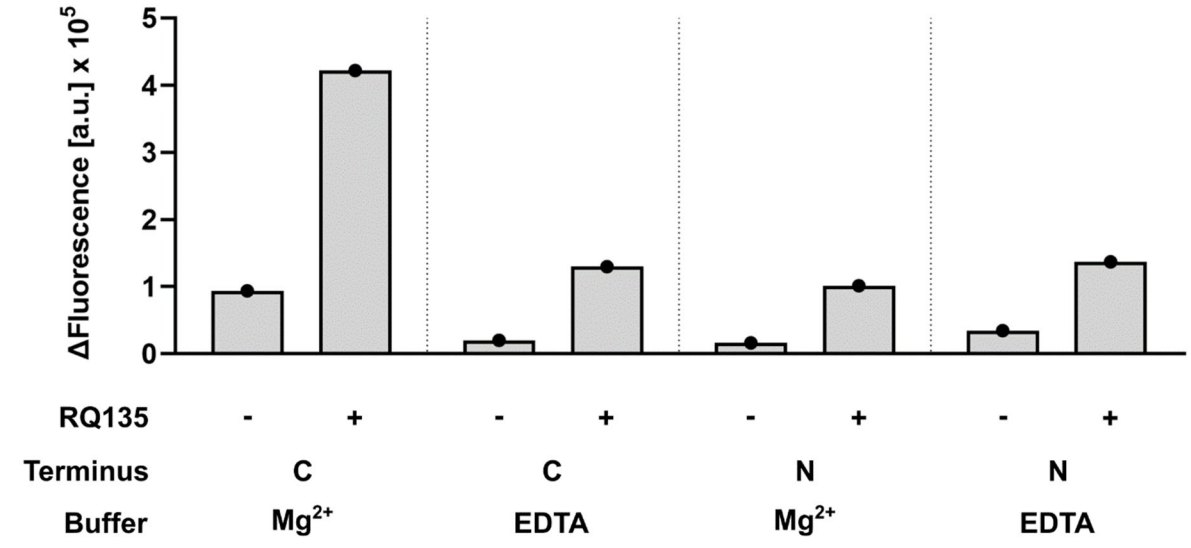
Supplementary Figure 6 Control digest of random insertion libraries: Primary DNA libraries were digested with either ApaLI or ApaLI and SspI. ApaLI cleaves twice, once directly before and once directly after the region coding for the MS2 genome. SspI cleaves only once, inside the antibiotic resistance gene (*kanR* or *camR*) of the inserted transposon. Cleavage of the plasmids with ApaLI generates two fragments, corresponding to the MS2 part (3.6 kb without insert, 4.8 kb with insert) and the plasmid backbone (1.2 kb). Cleavage of the plasmids with a combination of ApaLI and SspI generates two to three fragments, with only the backbone (1.2 kb) and the uncleaved MS2 part without insert (3.6 kb) producing distinct bands. Fragments of the MS2 part with an insert are cleaved by SspI into two fragments, generating a smear pattern. As the insertion occurs at random sites, fragments will have highly heterogenous size distribution with the higher molecular weight fragments being more visible.



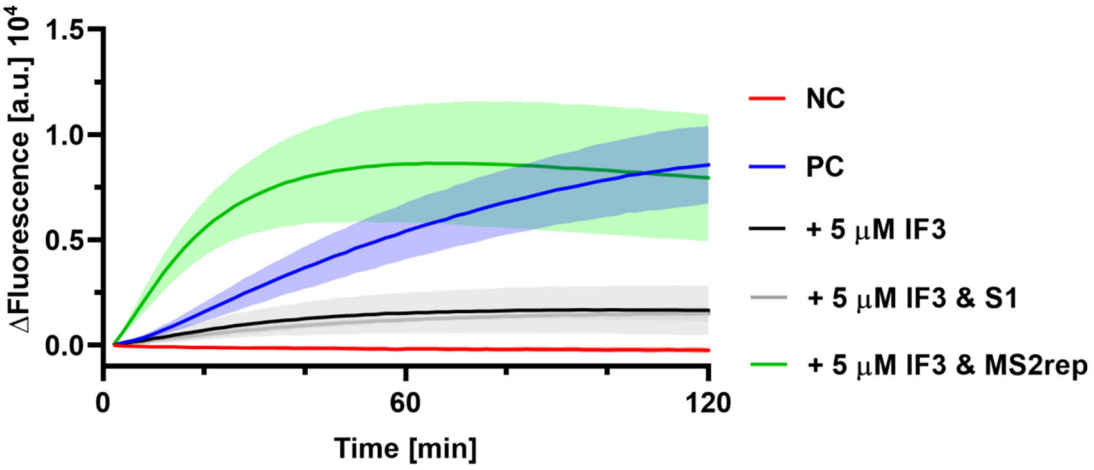
Supplementary Figure 7 Comparison of replicase preparations: SDS-Page of samples of MS2rep (1, 2) and Qβrep (3-6), respectively. Preparations differ in the position of the His_6-tag (N-terminal: 2, 5, 6 / C-terminal: 1, 3, 4), as well as buffer composition (Mg^{2+}: 1-3, 5 / EDTA: 4, 6). Red asterisks indicate the bands corresponding to the replicase subunits.



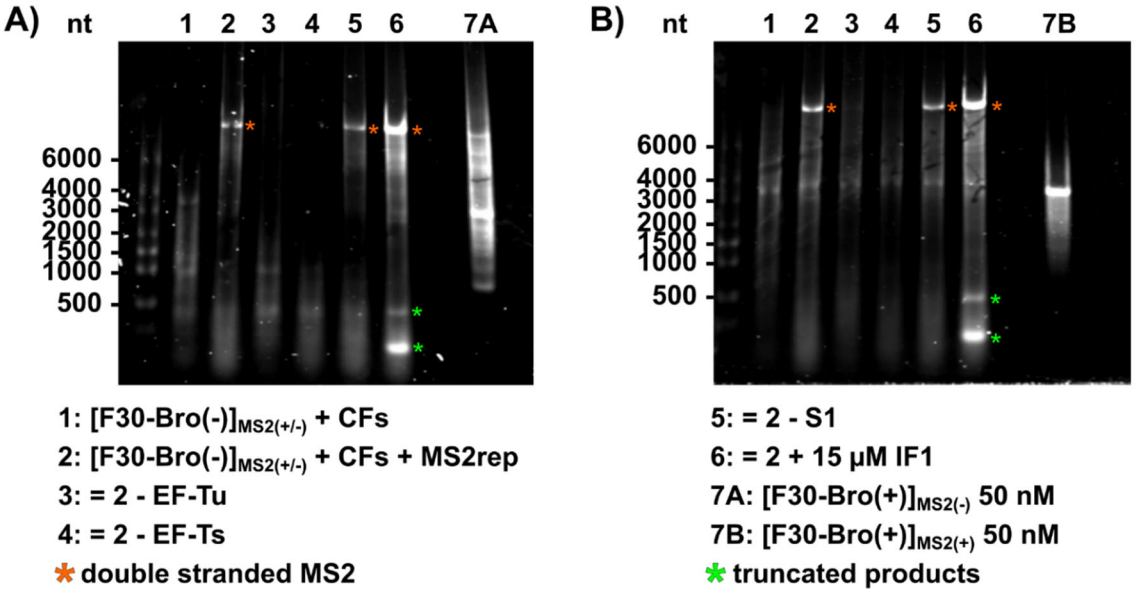
Supplementary Figure 8 SDS-PAGE of purified MS2rep co-factors and LD2 components: Samples for SDS-PAGE were taken after IMAC purification of the individual proteins. Loadout: BlueClassic Prestained Protein Marker (Jena Bioscience), EF-Tu, EF-Ts, IF1, IF3, MTF, AlaRS, AsnRS, IleRS, PheRS(α+β).



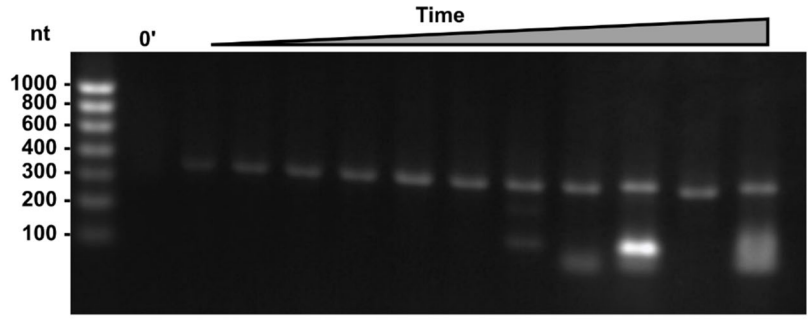
Supplementary Figure 9 Comparison of activity between different Qβrep preparations: Endpoint fluorescence for *in vitro* replication by different preparations of Qβrep, with (+) and without (-) input of RQ135 RNA. The preparations correspond to those mentioned in Supplementary Figure 7. Replicases were tagged with a His_6-tag at the C-terminus (C) or the N-terminus (N) and purified using buffers containing $MgCl_2$ (Mg^{2+}) or EDTA (EDTA). The reactions were all prepared by mixing the respective Qβrep (20 nM) with NTPs (1 mM each) in a buffer containing 10 mM $MgCl_2$, 100 mM Tris pH 7.5, 0.5 mM EDTA and 0.1 % Triton X-100, and incubated at 37°C for 75 minutes.



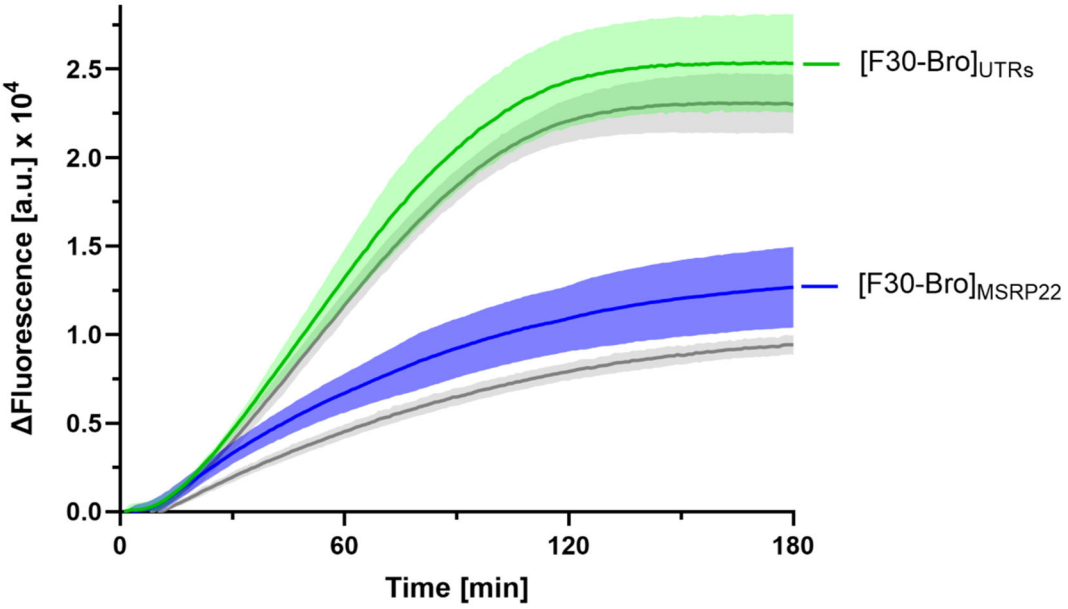
Supplementary Figure 10 Replicase inhibition by IF3 and rescue: Time resolved change of fluorescence intensity for *in vitro* replication reactions programmed with 50 nM [F30-Bro(-)]$_{UTRs(-)}$ without MS2rep (NC, red) or with 0.3 µM MS2rep (PC, blue), as well as with IF3 (black, 5 µM), IF3 with extra S1 (grey, both 5µM), or respectively, IF3 with extra MS2rep (green, both 5µM). All reactions also contained 1.5 µM S1, 15 µM EF-Tu and EF-Ts, and were incubated at 37°C for 2 hours. Error bars represent the standard deviation derived from three independent technical replicates.



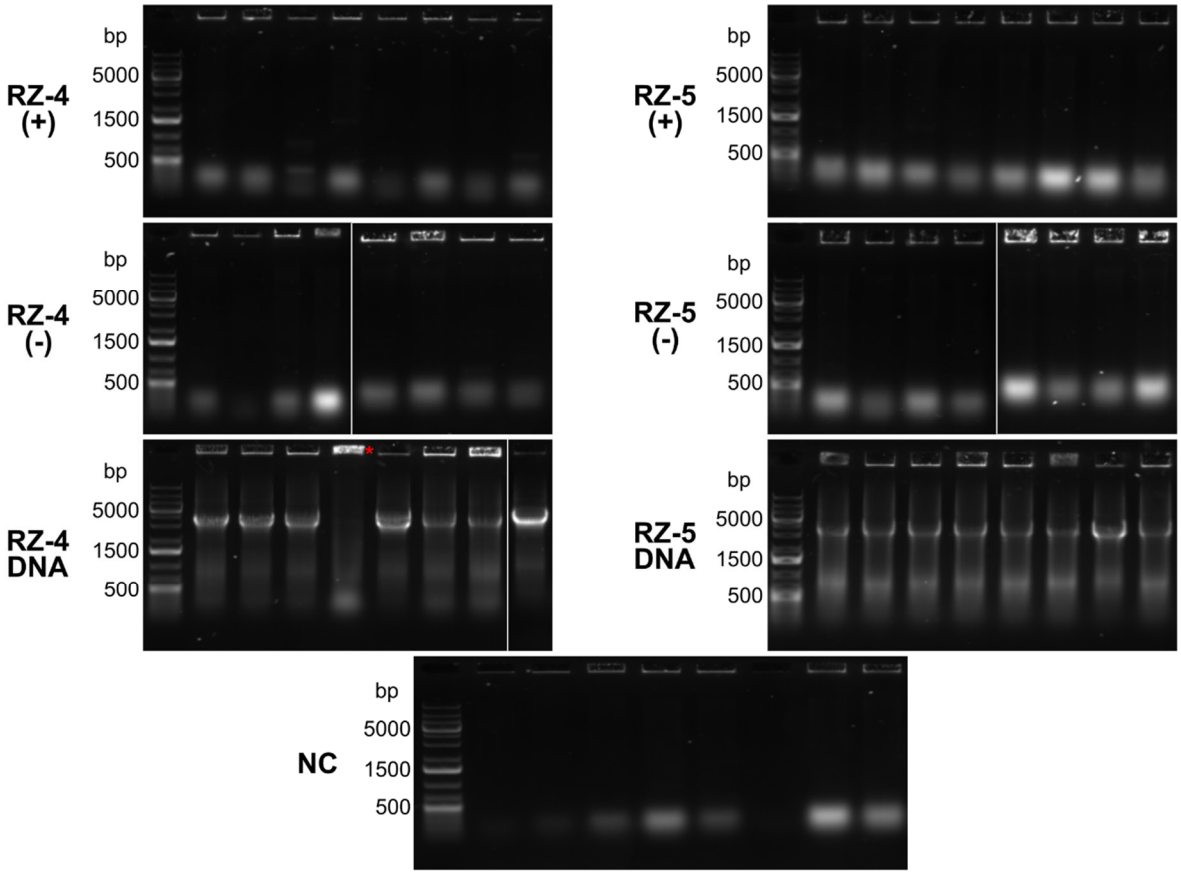
Supplementary Figure 11 *In vitro* replication of full length MS2: A) and **B)** Gel electrophoresis on 1 % TAE agarose gels of samples taken from the *in vitro* replication assays of full length MS2 RNAs [F30-Bro(-)]$_{MS2(+)}$ (**A)**) and [F30-Bro(-)]$_{MS2(-)}$ (**B)**), shown in Figure 24C and D, respectively. The samples were taken after incubation at 37°C for six hours, except for 7A and 7B, which are just size standards of the respective expected products [F30-Bro(+)]$_{MS2(-)}$ and [F30-Bro(+)]$_{MS2(+)}$ (both 50 nM) and were not incubated at 37°C. The reactions were supplemented with 50 nM of the respective RNAs, 1 µM MS2rep and the core host factors (CF) EF-Ts and EF-Tu and S1 in 15, 15 and 1.5 µM, respectively, if not indicated otherwise. IF1 was added in 15 µM concentration (6). Orange asterisks indicate double stranded [F30-Bro]$_{MS2}$ RNA, green asterisks indicate truncated products of the replication or potential small replicating RNAs.



Supplementary Figure 12 *In vitro* replication of [F30-Bro]$_{MSRP-1.0}$: Gel electrophoresis (1.5 % TAE agarose) of samples taken from an *in vitro* replication experiment of [F30-Bro]$_{MSRP-1.0}$, after incubation at 37°C for 0 / 5 / 15 / 20 / 25 / 30 / 35 / 40 / 45 / 50 / 55 minutes. The reaction mix was assembled with 250 nM MS2rep, 15 µM EF-Tu and EF-Ts, 1.5 µM S1 and programmed with 100 nM [F30-Bro]$_{MSRP-1.0}$ RNA (331 nt).



Supplementary Figure 13 Comparison of IVTxT of RZ-5 and [*rep*]$_{MSRP-1.0}$: IVTxT in PURExpress with RZ-1 as input mRNA with [F30-Bro]$_{UTRs}$ (green) and [F30-Bro]$_{MSRP-1.0}$ (blue) as read-out template. Reactions were prepared analogous to reactions for IVTxTs shown in Figure 29. Grey curves correspond to the respective IVTxTs with [rep(+)]$_{MSRP(+)-1.0}$ as input mRNA (Figure 29A, B). Error ranges indicate the standard deviation from three technical replicates.



Supplementary Figure 14 Colony PCR of cells transformed with RNA: Analysis of the source of Zeocin resistance by colony PCR of a subset of the colonies observed during the experiment depicted in Figure 31. Samples were analysed by gel electrophoresis on a 1 % TAE agarose gel. The red asterisk indicates a sample, where most of the DNA content got stuck in the well, presumably due to coprecipitation with cell debris. The size of the correct band for amplification from a DNA template is 3 kb, and this band was only observed for RZ-4(DNA) and RZ-5(DNA). Bands of lower molecular weight most likely arose from unspecific amplification.